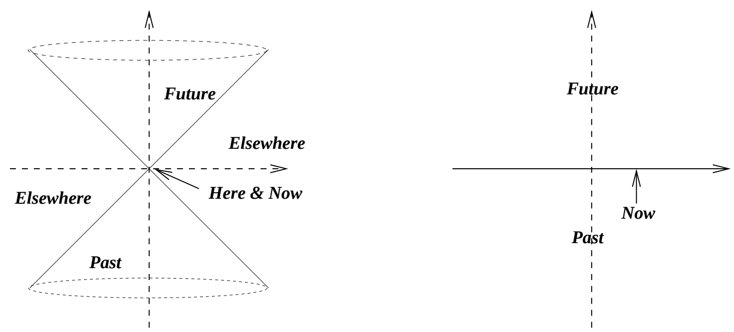
RELATIVITY
with a flashlight



(a) STR (b) Newtonian

A simple primer
using the Bondi K calculus

Ananda Dasgupta

March 3, 2007

Contents

Chapter 1

The background

1.1 Galileo - the father of relativity

1.2 Hunt for the elusive ether

1.3 Enter Einstein

Just what made Einstein embark upon his fantastic journey of discovery? There are many, many conflicting theories. Some would say that the Michelson-Morley experiment played an absolutely decisive role, some others would quote Einstein himself to show that he was unaware of the experiment in 1905. Einstein himself has often said that his training in electrical engineering as a polytechnic student had laid the seeds of enquiry about the deep significance of electromagnetism in his mind. However, we can be reasonably certain that whatever other influences played a role here, at one level the point that drove Einstein to his final theory was a philosophical one. The relativity principle as enunciated by Galileo appealed to him immensely - but he did have a very strong objection to it.

Galileo's principle applied to mechanics only - and this is what Einstein found hard to agree with. As he saw it, there is no such thing as a purely mechanical experiment - if nothing else you will at least have to take use light (and hence electromagnetism) to at least see the result! As far as Einstein is concerned, the principle of relativity was too beautiful to give up, yet, confined as it was too only mechanics, it was essentially an empty statement - devoid of any real physical meaning! Thus, the only way out, according to Einstein was to assume that the principle of relativity applied, not only to the laws of mechanics alone - but to all of physics! This principle became the basis of Einstein's special theory of relativity (STR), and today, when we speak of the principle of relativity, we mean
"All laws of physics are the same for all inertial observers."

This sweeping generalization of the principle of relativity from one confined to mechanics alone to one that covers all physical laws left Einstein with two options :

- The laws of mechanics, which are already known to be correct for all inertial frames need not be tampered with. This presupposes, of course, that the law of transformation when one goes over from one inertial observer to the other is the traditional, Galilean transformations. On the other hand, if the laws of electromagnetism are the same for all observers, then they can't be Maxwell's equations - since these are not Galilean covariant. Thus, the laws of electrodynamics must be modified.

- The laws of electromagnetism are correct - and they are the same in all inertial frames! This means that the transformation laws between inertial frames must be something other that the Galilean transformations. This also means that the

> laws of mechanics may fail to be covariant under the new transformations, and hence may require modification.

Faced with such choices I am pretty sure that most people would have sided with the first option. Electrodynamics was a fledgeling science at that stage. Mechanics, on the other hand, was the grand old man of physics - one, moreover, that had held sway over the way that scientists thought about the natural world for more than three centuries at that time. It was Einstein's genius that he did not hesitate to go the other way!

Chapter 2

The radar method

2.1 A few puzzles

Let's start with a simple puzzle. Suppose that I am walking down a long, straight road. There is a lamppost ahead of me - and I am walking straight towards it. A friend of mine crosses me on a bike, gets to the lamppost before he realizes that he has to tell me something important and then turns back to finally reach me. I have been walking steadily all the while. The question is - with respect to me, which half of his journey was longer, the outward one (from me to the lamppost) or the inward one (back from the lamppost to me)?

Although I am not a gambling man, I am ready to bet that most of you will say - "the outward one, of course!" Since I kept on moving all the while - it stands to reason I am much closer to the lamppost when he finished up, than when he started out. This answer is the expected one, but it is also *wrong!!!* Note that the question said, "*with respect to me*" - the answer that seems obvious to all of you is the one "with respect to the *ground*".

Try to look at the whole affair from my point of view. I am not moving with respect to myself - all that I see is the lamppost move steadily towards me. My friend crosses me, goes out - and meets the oncoming lamppost at some point. What distance has he covered with respect to me? From my origin to that point. Where does he turn back from? From that very point! Where does he land up finally? To my origin! Hence, with respect to me *he has covered the same distance on his outward journey as on his inward one*![1]

Let me stress once again that I did expect most of you to give the wrong answer. That is because despite learning a lot about relative motion, all of us instinctively think in terms of one observer all the time - the ground[2]! In fact, that is why most of you will be concerned with how far the lamppost is from me - the lamppost is fixed on the ground! That my friend travelled more during his outward journey is the correct answer - but only with respect to the ground! With respect to me, it is fifty-fifty.

> At this point, I can't refrain from airing a pet peeve of mine. I feel that the chapter on "relative velocity" that you study in high school does more harm than good! Most of the problems that are tackled under this topic tend to be like "The true velocity of rain is so and so, a man is moving with such and such a velocity - find out the apparent speed of the rain as seen by the man" or variants thereof. The trouble is - such language tacitly keeps on reinforcing the concept that there is such a thing as true velocity! Indeed, heading a chapter "relative velocity" immediately implies that relative velocities

---

[1]It is true that from my viewpoint, the lamppost is much closer to me when my friend finally reaches me, than when he started out - but that is not our concern!

[2]More precisely, it is the observer who is fixed with respect to the ground. I will take the linguistic liberty of calling the "observer fixed with respect to the object A" simply "the observer A", however - it cuts down the verbiage quite a lot!

> are somewhat different from ordinary velocity - where it should really be stressed that all velocities, by there very nature, are relative. It is nonsense to speak of a velocity (or acceleration, or even displacement, for that matter) without reference to some observer or the other! This is simply because all of these quantities are related to the position - and the position *must* be measured with respect to an observer. The so called “true velocity” almost invariably turns out to be the velocity with respect to the ground - this only keeps on reinforcing the instinctive idea that there is something special about the ground as an observer.

Let’s come back to our puzzles. The next one is - if the speed of my friend on the bike was the same on both halves of the journey, then which half took him more time? This one is really simple - since the outward journey was longer than the inward one - the first half must have taken more time.

“Hold on!” - some of you must be screaming at this point. This one is obviously correct with respect to the ground. Since the two journeys are equally long with respect to me, shouldn’t the times be equal when I measure them? If this is your answer, then congratulations! Not because your answer is right (it isn’t!) but because you have unwittingly surmounted the biggest stumbling block that lies in the path of understanding the special theory of relativity! Jokes apart, think about this issue a bit - would you really expect the times to change depending on whether a man fixed on the ground sees the events or whether I, who am moving, sees it?

The situation, once again, is very simple. It is true that the distances my friend covers in the two halves are the same with respect to me. The speeds are also the same - but with respect to the ground! When he goes away from me, his speed, as I see it, is less (it is $v - u$ if v and u are the bike's and my speed with respect to the ground, respectively) than when he is coming back in (it is $v + u$). So, he takes more time to go out than to come back in - just as the ground sees it. Indeed, if you take the trouble of calculating the times from my viewpoint (you should!) you will see that they exactly match the time that the ground observes[3]. The moral of the story is - when you are using somebody's displacement measurements, you should not mix them up with someone else's velocity measurements!

Brace yourself - for this is where our hero the flashlight makes its grand entrance into the story! Consider the same scenario as before - except that instead of the lamppost I am now moving towards a mirror fixed to the ground. I shine a flashlight at the mirror. A pulse of light travels to the mirror and comes back to me. As with my friend the biker, the flash of light travelled more during the outward journey than during the inward one - *with respect to the ground.* Again, both halves of its journey were equally long with respect to me. What about the times? It stands to reason that the outward journey takes more time than the inward one for a ground based observer. What do I say about this? Now, as per Einstein, light moved with the same speed c for both halves of its journey, not only with respect to the ground - but also with respect to me! This leaves me with no option but to conclude that - with respect to me light takes the same time to complete the two halves of its journey! *Accept the constancy of the speed of light, and you have to accept that time depends on the observer*!

Our simple little puzzle has led us right into the heart of the

[3]This should not come as a surprise. After all, this is how we arrive at the $v \pm u$ results in the first place!

special theory of relativity. As we will see, time and again - this dependence of time on the observer is the major point of departure that makes STR distinct from Newtonian physics. To stress this difference a bit further, let me put this in a slightly different way.

Newton would be quite comfortable with the idea that two events that occur at the same place but at different times according to one observer, may happen at different places according to another. He would certainly have a fit, however, if someone were to suggest that two events that happened at different places but at the same time according to me - happens at different times according to someone else!

Just think of a diner in a railway car starting on his soup as the first event and finishing up the last spoonful of dessert as the second one[4]. To the waiter in the dining car, both of these events occur at the same place - the dining car table. For someone on the ground, however, the two events may have occurred very far apart indeed - the distance the train has travelled while the diner was busy with his food! However, if the waiter sees another traveller take a sip of water at the other end of the dining car at the same instant when our diner is paying the bill - our instincts (which form the basis of Newtonian physics) will tell us that someone on the ground will also agree that the two events are simultaneous. STR puts space and time on a much more equal footing - though. It actually tells us that two events that occur at the same time but at different places according to one observer, may appear to occur at different times to another!

This leaves us in a somewhat difficult position. Newtonian physics has the advantage of a universal time which serves as a fixed back-

[4]My apologies to George Gamow and Cleveland for lifting this example almost verbatim from their classic text - *Physics - Foundations and Frontiers.*

ground that keeps a strict order on things. In Newtonian physics, the issue of measuring time was more of a matter of technology - all you had to do to keep time better was build a better clock! The deeper issue of exactly what measuring time means never reared its ugly head there. However, with a time that changes from observer to observer, STR forces us to take a deeper look at the way in which we measure space and time. Indeed, the reason why STR is fundamentally important is the fact that it forces us to take a deeper look at the very nature of space and time itself.

2.2 The Space-time diagram

For convenience I will now pretend that the world has only one space dimension. Most of the amazing new things that relativity throws at us makes their appearance in 1 dimension - the transition to three space dimensions from one is mostly devoid of new shocks. Don't worry, I will indicate the three dimensional generalization of our results as and when needed.

Let me introduce a device that will be very useful to us in the long run. This is none other than your familiar x vs. t diagram - but with a slight twist. In this diagram I will plot x along the *horizontal* axis and t along the *vertical* axis. Moreover, I will be plotting not t, but ct along the vertical axis. This makes it possible for me to refer to both the directions using the same units. You could also say that I have chosen a system of units in which space and time are measured by the same units (meters in this case), where a meter of time is the time it takes light to travel 1 meter (this is about 3.3 nanoseconds in our conventional units). If I adopt this line, then velocity becomes dimensionless! In particular, the speed of light in this new system of units is just - *one*.

Figure 2.1 shows a typical space time diagram. In this diagram, the path of light waves will be straight lines inclined at an angle of $45°$ to the axes. The solid red line is that of a beam of light emitted by the observers flashlight at $t = 0$. The dashed red line is also that of a light beam - but one travelling to the left. Moving particles trace out a set of points in this diagram - the line joining them is called the **worldline** of the particle. Like in our familiar $x - t$ diagrams, the lines will be straight for particles moving with a uniform velocity and curved for an accelerating one. One word of warning - since we have reversed the positions of time and space axes, the slope of a worldline is the reciprocal of the particle's velocity (in units of c) rather than the usual velocity. Since one of the results that follow from relativity (one that I will prove later) is that material particles can't travel faster than light, all worldlines, straight or curved, must always be steeper than lightlines - *i.e.* make more than a $45°$ angle with the x axis. The blue line in the figure represents the worldline of a particle moving with uniform speed, while the brown one is that of an accelerated particle. Both the green lines are those of "impossible" particles - the straight one depicts a particles moving uniformly faster than light, while the curved one is that of a particle that moves slower than light at the beginning and end of its track - but is faster than light over the flat middle portion. Finally, what about the worldline of the observer herself? Since she is fixed at $x = 0$, her worldline goes straight up in the diagram - it is the t axis!

Space-time diagrams like this one are very useful in understanding various concepts of relativity. Let me stress, though, that the reason why I was able to depict our spacetime on a diagram is because I decided to ignore two of the spatial directions. If I had tried to do this for full three dimensional space, then I would have to use up all three directions for the spatial axes - which will leave me nowhere to put the fourth, time, axis! With one space dimension only, spacetime is two dimensional - something that I can draw very easily on a sheet of paper.

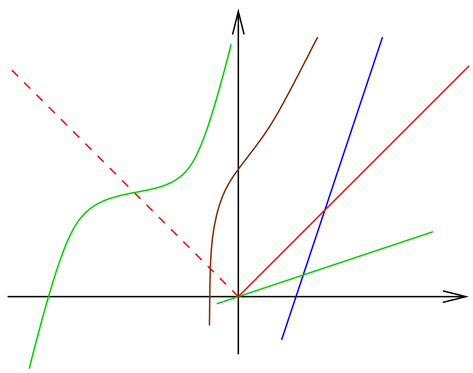


Figure 2.1: A space time diagram.

Although we cannot draw a four dimensional spacetime diagram on a sheet of paper - nothing prevents us from understanding it at a mathematical level, or even imagining such a thing. There is nothing magical about four dimensions - all that we are saying is that events must be labelled by four numbers - the four coordinates x, y, z and t. Of course, this was true even in Newton's world. There is a deeper sense, though, behind the statement that spacetime is four dimensional in STR - I hope to make this clear in a coming section.

There is one thing about four dimensional spacetime that I must bring up here. In the 2D space time in figure 2.1, the light lines that go through any point form a pair of straight lines, both inclined at $45°$ to the two axes. In four dimensions, the equation that describes light emitted at $x = 0$ and $t = 0$ obeys the equation

$$c^2t^2 - x^2 - y^2 - z^2 = 0 \tag{2.1}$$

- which is simply the statement that light travels equally in all directions with the uniform speed c. Although we can't draw the surface corresponding to this equation, we can definitely understand its content. What I can draw, is the version of (2.1) in which I use only two space dimensions, ignoring z. This equation, $c^2t^2 - x^2 - y^2 = 0$, is that of a cone. This gives rise to one of the most important pieces of jargon in relativity - the surface described by (2.1) is called - the **lightcone**. I will have a lot more to say about the lightcone later.

2.3 Measuring space-time with the flashlight

A simple way of measuring distances and times that characterize where and when an event occurs follows simply from the discussion in section 2.1. All I have to do is to shine our flashlight and wait for the light to bounce back! Actually I have to do a bit more - I must note the times at which the flash was sent out and came back in - both according to my wristwatch. This is shown in figure 2.2. Since light must have travelled the same distance going out from me to the point where it bounced and back to me and must have covered both of the halves at the same pace - the bounce must have occurred just midway in between. If these times are t_1 and t_2, respectively, then the time and position of the bounce must be

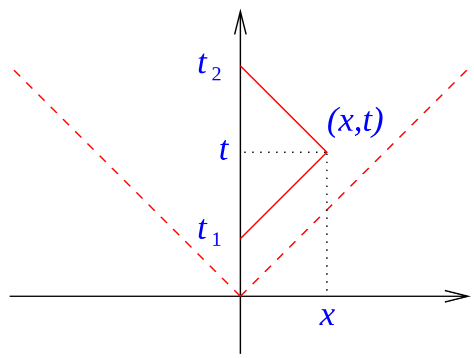


Figure 2.2: Measuring space and time with a flashlight

$$t = \frac{1}{2}(t_2 + t_1) \quad \text{(2.2-a)}$$

$$x = \frac{c}{2}(t_2 - t_1) \quad \text{(2.2-b)}$$

You may find this hard to swallow - but almost all of relativity can be derived from this nearly trivial set of rules!

Our humble flashlight is in rather good company - the all important radar systems that help armies locate and shoot down enemy planes and air traffic controllers guide friendly ones to a safe landing works on exactly the same principle outlined above! So, I will refer to this way of measuring space and time as the **radar method**.

The rules that I have laid down above may seem too simple to be true. More importantly - they seem so much in tune with our classical notions that it is hard to see just how the conclusions that we can draw from these rules can differ from the classical ones. Let me point out just where the rules cease to be classical. In Newtonian mechanics, the equations (2.2-a) and (2.2-b) will be valid, but only in the one unique frame where the speed of light is the same in all directions. This is the frame fixed to the medium in which light propagates - the luminiferous ether. Einstein freed physics from the shackles of the ether - so that in STR (2.2-a) and (2.2-b) are valid for all inertial observers. Thus, if an observer is moving away from me, he can measure the in and out times of a light beam and use the same equations to locate the bounce in space and time - but this time according to his own coordinates.

One of the major new features in STR is that space time measurements change from observer to observer in a rather new, unexpected manner. Since all inertial observers can play the game of measuring spacetime coordinates by using the equations (2.2-a) and (2.2-b), we can find out how these changes work out by playing with the rules of this game. This is precisely what I intend to do in the next few sections.

2.4 The K factor

Let us now consider two players in the game of measuring space-time coordinates. I will call them Alice and Bob. Bob crosses Alice just when Alice's watch reads zero and he also synchronizes his watch to read zero at the same time. From Alice's viewpoint, Bob moves to the right with a uniform speed v.

Both Alice and Bob are carrying flashlights of their own. Alice keeps shining her flashlight in Bob's direction at a regular interval of τ. If Bob had been standing still, he would have received the flashes of light the same interval apart. Since he is moving away from Alice, each successive flash of light takes longer to catch up with him - so he receives the signals at a larger interval. Let's call the factor by which this interval is larger K. So the flashes of light will reach Bob at intervals of $K\tau$, as read by the watch carried by him. This last point is of crucial importance - remember that I have already shown you that the watches of Alice and Bob may not agree.

The K factor will play central role in the story of how Alice and Bob's readings of events are related to each other. It is called the Bondi K factor in honour of Hermann Bondi - who introduced this delightful little character in the stage of relativity theory.

Think about it a bit, and you will realize that you have already seen the K factor in action. Alice did not really have to shine the flashlight on and off to produce a periodic signal - if she had just kept on shining the flashlight steadily the electric field in the light will go up and down - reaching a maximum once every time period. Bob will receive the peaks at a larger time interval - thus the time period that Bob sees is greater by a factor of K. So, Bob sees the light to have a smaller frequency that Alice does. This is nothing other than Doppler effect - remember how the whistle of a train

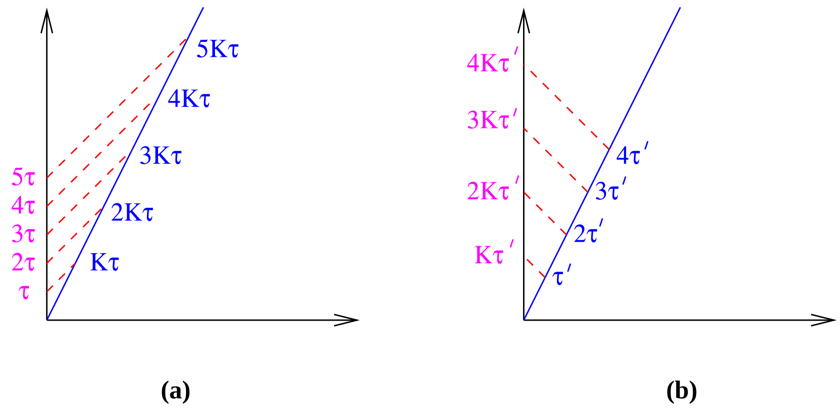


Figure 2.3: The K factor. In (a) Alice sends Bob flashlight signals at an interval of τ, which bob receives at an interval $K\tau$. (b) Shows what happens when Bob sends his signals - note that the same factor K shows up again. Note that the times shown in purple are times read by Alice's watch, while those in blue are readings from Bob's watch.

engine suddenly drops in pitch when the engine crosses you? This is why another name for K is the Doppler factor.

What happens if Bob decides to shine his flashlight at Alice? If he flashes his flashlight at an interval τ' as measured by his watch, Alice will receive them at a larger interval of time according to her watch. What may come as a surprise to you is that the interval that Alice will see is exactly $K\tau'$. The same factor K comes in, both for Bob receiving Alice's signals and for Alice receiving Bob's signals!

Why are the two factors the same? It is the basic principle of relativity, the fact that all inertial observers are exactly equivalent, in action! Note that if Bob had moved away from Alice to the left instead of to the right, he would still have received Alice's signals at the same interval as long as his speed is the same as before - there being no difference between left and right. Now, with respect to Bob, Alice moves to the left with the same speed to the left as

Bob is moving to the right with respect to her. Thus any difference between the two factors will point to a fundamental difference between Alice and Bob as observers. This precisely is what the principle of relativity rules out! Imagine that Alice has flashed her flashlight at time τ according to her wristwatch at Bob, who receives it at a time when his watch reads $K\tau$. Instead of letting the light beam flash past him, Bob holds up a mirror that reflects the pulse back to Alice. This beam will reach Alice at $K^2\tau$, who reflects it back to reach Bob at $K^3\tau$, and so on ... Figure 2.4 shows this bouncing to and fro of the flashlight beam.

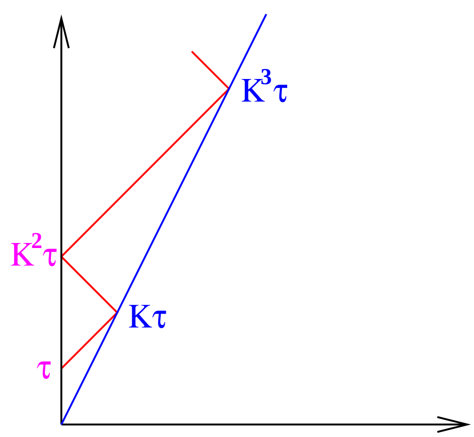


Figure 2.4: Determination of K

When and where does the first bounce at Bob's mirror happen, according to Alice? Our basic rules of spacetime measurement, (2.2-a) and (2.2-b) tells us the answer :

$$t = \frac{K^2+1}{2}\tau$$
$$x = c\frac{K^2-1}{2}\tau$$

One thing must be immediately clear - the time at which Bob saw the first bounce occur was at time $K\tau$ - which is clearly different from the time $\frac{K^2+1}{2}\tau$ that Alice infers for it.

To find the value of K, just notice that

$$\frac{x}{t} = \frac{K^2-1}{K^2+1}c$$

Since x is the position where Bob is at time t, according to Alice, the ratio $\frac{x}{t}$ is nothing but the speed v of Bob that Alice sees. So, calling the dimensionless quantity $\frac{v}{c}$ β, the equation translates to

$$\beta = \frac{K^2 - 1}{K^2 + 1} \tag{2.3}$$

which can be easily solved to find K

$$K = \sqrt{\frac{1+\beta}{1-\beta}}. \tag{2.4}$$

Thus the K factor goes from 1 when Bob is stationary with respect to Alice, to ∞ when Bob's speed approaches the speed of light $(\beta = 1)$. The factor is smaller than 1 for negative values of β, and falls all the way to 0 for an observer moving to the left with a speed c. Indeed, reversing the sign of the velocity changes the K factor to its reciprocal.

If Bob moves away from Alice with a speed v, then Alice must move away from Bob with a speed $-v$. It follows, then, that the K factor of Alice with respect to Bob must be K^{-1}, if that of Bob with respect to Alice is K. However, isn't this contrary to what I said a while ago - that when Bob sends here signals at time intervals of τ, Alice must receive them at intervals of $K\tau$ (and not $K^{-1}\tau$), according to the principle of relativity?

If you think about this a bit, you will realize that there is no contradiction at all! Imagine a third fellow, maybe Charlie, who is situated well to the left of both Alice and Bob, who is also shining flashes of light at both of them. Both Alice and Bob sees the flashes travelling to the *right*[5] with speed c. If Bob sees these flashes at a

[5]That the direction as well as the speed of light stays unchanged when you jump from one inertial frame to another is a consequence of the fact that "you just can't outrun light"! I will show you how this result follows from the relativistic rules of velocity addition in section 3.4.

time interval of τ, then Alice must see them at the smaller interval $K^{-1}\tau$. This exactly, is the meaning of the statement that the Doppler factor of Alice with respect to Bob is K^{-1}! In this case, Alice is moving, as far as Bob is concerned, in the direction of the source of the light and is hence getting to meet each successive pulses sooner than Bob does. When Bob sends Alice the signals, though, the light signals must move towards the *left* to get to the Alice, which is the same direction that Bob sees Alice move away in! The light signals take some extra time to catch up with the receding Alice - hence, the increased time interval of $K\tau$.

In summary, then, the Doppler factor K for Bob with respect to Alice is what you must multiply time intervals between light signals as seen by Alice, to get that seen by Bob - but for light travelling towards the positive x direction, namely, to the right! For light travelling towards the left, the multiplying factor is not K at all, but rather K^{-1}!

In terms of frequencies, Bob must divide the frequency of light that Alice sees by K to get the frequency that he sees, but again, only for light travelling to the right. For light moving the other way, Bob will have to multiply by K, instead of dividing by it!

Let me address another point that you may have already thought of on your own. The equation (2.4) clearly shows that the value of K ceases to be real once β exceeds 1 in magnitude. Does this prove the statement that I have made all along - that it is impossible for a material object to move faster than light? Unfortunately, the answer is - no! Just think about it a bit and you will realise that all it means is that if Bob were to recede from Alice faster than light, then the flashlight signals Alice sends out towards him will *never* catch up with him! To understand why material particles cannot exceed the speed of light, you will have to wait a while more!

You may be inclined to think of (2.4) as more physically significant than (2.3). After all, velocity is something we all know and understand - while the K factor is something of an unknown quantity. However, now that I have shown you that measurements of space and time may not be as straightforward as it may have seemed, it should not come as much of a surprise that the speed is not that easy a quantity to measure directly. On the other hand, K is something that can be measured directly with a lot of ease. All Alice and Bob have to do is measure the frequency of a given flash of light. The ratio of these values gives us the K factor directly. So, in a practical situation, it is more likely that we will calculate β from K rather than the other way around. Indeed Alice may just as well describe the motion of Bob by his K factor rather than his speed - and this will turn out to be quite convenient for many purposes.

> In order to stress the difference between the STR scenario and the classical one, let me show you what the K factor would have been in a world that obeyed Newton's physics. The first light pulse, sent at $t = 0$, reaches Bob immediately, since he is right next to Alice at this point. If Bob receives the second flash at time τ' (note that this is classical physics - we are not bothered about who measures this time), then the second flash must have travelled for a time of $\tau' - \tau$. This gives $c(\tau' - \tau) = v\tau'$, leading to
>
> $$K = \frac{\tau'}{\tau} = \frac{c}{c - v} = \frac{1}{1 - \beta}$$
>
> What about the signals Bob emits? Again, Bob's first signal reaches Alice immediately. When Bob emits the second flash, at a time $\bar{\tau}$ (say), he is at a distance of $v\bar{\tau}$ from Alice. The flash covers the distance back to Alice in

a time given by $\frac{v}{c}\overline{\tau} = \beta\overline{\tau}$, so that Alice receives the flash at $(1+\beta)\,\overline{\tau}$. So, the K factor for Alice for signals emitted by Bob turns out to be

$$K' = 1 + \beta$$

Note that unlike STR's prediction, the two K factors are different! Obviously, both differ from the STR prediction, (2.4). Note, though, that since $(1+\beta)^{-1} \approx 1 - \beta$ for $\beta \ll 1$, both these factors are quite close to the relativistic value for low speeds.

Why do the two K factors differ? If you carefully go through the derivation above, you will see that despite appearances, there is a big difference between Alice and Bob as observers. I have tacitly assumed that the speed of light is c with respect to Alice, *i.e.* she is fixed with respect to the ether! In the classical picture, then, Alice does have a privileged position! Indeed, note that in the propagation of sound, it does matter whether the source moves or the observer moves, because in each case there is a preferred reference frame - one fixed to air, the medium of sound propagation.

Chapter 3

Kinematics from the flashlight - the K calculus in action

3.1 Time Dilation

Let's go back to figure 2.4 again. As you have already seen, Alice and bob disagrees about the the time at which a flash sent out by Alice at time τ meets Bob. According to Bob, this happens at a time $K\tau$, whereas Alice reckons this to be $\frac{K^2+1}{2}\tau$. It is pretty easy to sees that Bob's time is the smaller of the two, no matter what the value of K is[1]!

You may object that this destroys the equivalence of the two players in this game. After all, if Bob always measures the smaller time, then he can not be on the same footing as Alice! Note, however, that Bob does have a special role in the event that we are

[1] $\frac{K^2+1}{2}\tau - K\tau = \frac{1}{2}\left(K-1\right)^2\tau > 0$ - for all real values of K. Note that here (as well as elsewhere) we are assuming that Bob can not move away from Alice at a rate faster than light.

talking about - *he is on the spot*! This means that he had the luxury of directly using his watch to time the bounce, while poor Alice had to be content with an indirect determination of the time. Indeed, if we look at the second bounce the light flash suffers, we see that Alice times it at $K^2\tau$, while Bob *infers* that it occurs midway in between the first and third bounces, *i.e.* $\frac{K^3+K}{2}\tau$. Once again, the "on the spot" time (this time its Alice's turn) is smaller than the inferred time. There is nothing special, then, about either Alice or Bob - whoever happens to be on the spot measures a smaller time.

So, STR does make a distinction between a time measurement made directly and one made indirectly. In STR jargon, directly measured time is called the **proper time**. Going by this, you may be tempted to call the indirectly measured time the "improper time" (some authors do). In my opinion, this latter term is not very suitable - since it gives us the notion that something is wrong about such a time measurement!

Going back to the first bounce suffered by the flash of light in figure 2.4, we denote the proper time at which it occurs (Bob's time) by t_0 and Alice's time by plain t. We see that

$$\frac{t}{t_0} = \frac{K^2+1}{2K} = \frac{\frac{1+\beta}{1-\beta}+1}{2\sqrt{\frac{1+\beta}{1-\beta}}} = \frac{1}{\sqrt{1-\beta^2}}$$

which gives us the famous **time dilation** formula

$$t = \frac{t_0}{\sqrt{1-\beta^2}}. \tag{3.1}$$

The factor $\frac{1}{\sqrt{1-\beta^2}}$ recurs so often in STR it deserves a special symbol of its own, namely γ :

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} = \frac{K^2+1}{2K} \tag{3.2}$$

and so

$$t = \gamma t_0.$$

Note that the ratio of measured times is the same for the second bounce of the flash as well - but this time t is Bob's time, while t_0 is the time as measured (directly) by Alice.

Although equation (3.1) is pretty straightforward - it can lead you into all sorts of trouble, especially if you get confused about which observer's time is t_0 and which one is t. So, let me stress this once again, *if an observer is present at the location of both the events (on the spot) then the time interval he measures between them is the proper time.* If you are talking of two events such that neither Alice nor Bob is on the spot for *both* of them, then neither can claim his time measurement to be the proper time. How, then, are the two time's related to each other? Well, you will just have to wait till we get to the derivation of Lorentz transformations in a little while.

3.1.1 Moving clocks run slow!

Let's try looking at the watch Bob is carrying with him from Alice's point of view. Two successive tick on the watch are two events and both Bob and Alice can time the gap between them. Since Bob is present at both the ticks (its his wristwatch, after all!), he can directly read off the time interval. So, Bob measures the proper time interval between the ticks, while Alice does not. So, if Bob says that the gap between two successive ticks on his watch is one second, Alice reckons the gap to be larger - γ seconds to be precise.

From Alice's point of view, Bob's watch is running slow!

You must have caught on by now to the conclusion that as far as Bob is concerned, it is Alice's watch which is running slow - by the same factor γ. This leads to one of the most surprising claims of STR - **moving clocks run slow**! Does Bob notice his own clock run slow? Not on your life! That would have given him the ability to detect his own motion just looking at his watch - something that an inertial observer just can't do!

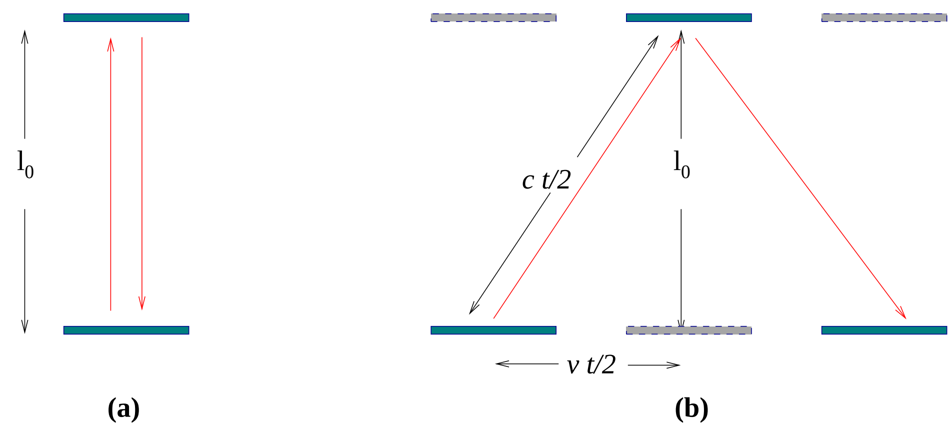


Figure 3.1: An idealized clock. (a) The time taken for light to return to the lower mirror is $\frac{2l_0}{c}$, (b) The time taken is longer for someone moving with respect to the clock, since light has to cover a bigger distance.

Let me show you a more direct proof of this effect of a moving clock running slow. To do this, you have to imagine an actual clock, albeit a highly idealized one! Imagine a contraption which has two mirrors, a distance l_0 apart. As shown in figure 3.1a, a ray of light leaves the lower mirror and bounces from the top mirror to return. This to an fro motion of light defines one period for our clock, which I will refer to later on as the "photon clock". If I stand still next to this clock, I will see light cover a net distance of $2l_0$ - which means the time period of the clock with respect to me is

$\frac{2l_0}{c}$. Note that since both the start and end of the light rays to and fro journey occurs in the same place with respect to me, this time interval is the proper time interval between the two events.

Consider what this will look like to someone else, who sees me and my clock move uniformly to the right with a speed v. He will see light leave the lower mirror, reach the upper one and bounce back to the lower one, just as I do. However, by the time light reaches the upper mirror, it would have shifted to the left ... - hence he will see light follow the path shown in figure 3.1b. The path is obviously longer than the one I see. If this had been Newtonian physics, he would have claimed that light is moving faster (at a speed $\sqrt{c^2+v^2}$ to be exact), so that it covers the path in the same time $\frac{2l_0}{c}$ that I see. That option is out - STR tells us that light has to make the trip at the same speed with respect to him too - and thus he must see the trip take a larger time. Thus one tick on my moving clock takes more time than the designated $\frac{2l_0}{c}$ as far as he is concerned - it is running slow!

To find out just how slow, all we have to do is appeal to good old Pythagoras! If t is the time taken for the round trip according to him, then light takes half of it to get to the upper mirror, covering a distance of $\frac{ct}{2}$. In this time, the upper mirror has moved a distance of $\frac{vt}{2}$. So, in figure 3.1b we have a right angled triangle with sides l_0 and $\frac{vt}{2}$ and a hypotenuse $\frac{ct}{2}$. So

$$\left(\frac{ct}{2}\right)^2 = l_0^2 + \left(\frac{vt}{2}\right)^2$$

which can be easily solved to give

$$t = \frac{\left(\frac{2l_0}{c}\right)}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{t_0}{\sqrt{1-\beta^2}}$$

- exactly as our time dilation formula tells us.

3.1.2 Accelerated clocks

So, a clock that is whizzing past you rapidly does run slow compared to your clock. But what if, instead of just moving at a steady pace, the clock also accelerates? This was a pretty thorny question in the formative years of relativity - but today we know the answer: acceleration does not matter! This does not mean that your real clock is unaffected by accelerations - drop one on the floor and it is likely to break! What it does mean is that it is in principle possible to build clocks that are not affeceted by accelerations and that in reality we can get pretty close to this ideal. So a nonuniformly moving clock whose speed is u slows down by the same factor of $\sqrt{1-u^2/c^2}$ as a uniformly moving one. The only trouble is that this factor varies with the speed of the particles and hence with the time. Thus the factor can be used, but only over infinitesimal periods of time:

$$d\tau = dt\sqrt{1-\frac{u^2(t)}{c^2}} \tag{3.3}$$

Here we have used the notation τ for the proper time as observed by an arbitrarily moving observer. We will reserve the notation t_0 for the case where the observer is moving uniformly. The net proper time that elapses between two events $\mathcal{A}$ and $\mathcal{B}$ is thus given by

$$\Delta\tau = \int_{\mathcal{A}}^{\mathcal{B}} dt\sqrt{1-\frac{u^2(t)}{c^2}} \tag{3.4}$$

Of course, this reduces to (3.1) whenthe velocity of the clock is a constant, $u(t) = v$.

3.1.3 The Doppler factor revisited

In section 2.4 I presented one of the major results of this book - the Doppler factor for a moving observer. I also showed you that the value calculated from STR differs from the classical result. Armed with our result for time dilation, we may now re-derive the formula (2.4) more directly.

As you have seen, the classical value for the K factor is $\frac{1}{1-\beta}$. This means that classically, a signal that Alice sends out at time τ would reach Bob at a time $\frac{\tau}{1-\beta}$. And so it does - even if we take STR into account! If you take a look at the derivation there, all I have used is the fact that in a time τ' , Bob has moved away from Alice by a distance of $v\tau'$ and at that instant the light pulse, which has been travelling for a time $\tau - \tau'$ must have moved a distance of $c(\tau - \tau')$. Equating these two gave me the time τ', the time when Alice's second pulse reaches Bob. That sort if impeccable reasoning is something you just can't fault! The only catch is that, while the second pulse that Alice sends out does reach Bob at a time given by $\tau' = \frac{\tau}{1-\beta}$, this is the time according to Alice's watch! What does Bob's watch read at this instant? Since his watch is running slow compared to Alice's by a factor of $\sqrt{1-\beta^2}$, his watch must be reading

$$\tau' \times \sqrt{1-\beta^2} = \frac{\tau}{1-\beta}\sqrt{1-\beta^2} = \tau\sqrt{\frac{1+\beta}{1-\beta}} \tag{3.5}$$

which gives the K factor according to (2.4)!

3.1.4 Testing relativity

Is time dilation for real? The first question that you may raise is, if moving clocks really do run slow, how come no one had noticed? The answer should be immediately obvious! All the clocks that we see around us moves too slowly for the effect to show up! Consider a clock that rushes past you on an aircraft flying at Mach 1 (the speed of sound). How much does this clock slow down by? To get an estimate of this, note that the speed of sound is about 300 m s^{-1}, while that of light is 3×10^8 m s^{-1}, so that the β for our clock is 10^{-6}. So, this clock goes slow compared to the one that sits still next to you by a factor γ given by

$$\gamma = \left(1 - 10^{-12}\right)^{-\frac{1}{2}} \approx 1 + 5.0 \times 10^{-13}$$

So, this clock that flies past you does go slower than yours - but only by five parts in 10^{13}! No wonder no one had noticed!

Having said that, I must hasten to add that today's technology is good enough to actually detect even such a small effect! Today, we have many direct verifications of the time dilation formula (3.1) - something which would have been unthinkable in Einstein's time.

What's more, the time dilation formula is no longer a mere curiosity producing either tiny effects that would not be detectable except with very sophisticated technology, or large effects for exotic objects that travel very fast - it is almost a part of daily life today! You may have heard of the Global Positioning System (GPS) - a network of satellites that receives signals from transmitters attached to vehicles (like ships in the open sea, aircrafts or even your mobile phone!), times them and uses something very much like high school trigonometry to triangulate the exact location of the source, to within a few meters. The clocks on the satellites run slow compared to the clocks on the sources (because of their large orbital speeds) and you must compensate for this effect to achieve this accuracy! Without this compensation, GPS calculations would have been off by miles! This perhaps is one of the most drastic confirmation of STR[2] that affect our daily lives.

3.1.5 The twin paradox

Scientists love paradoxes - apparently impeccable arguments that lead to absurd or contradictory results. Resolving paradoxes helps them hone their logical skills, and sometimes even make new discoveries. Perhaps no other branch of physics has given rise to so many paradoxes as the special theory of relativity. One of the best known is the notorious twin paradox - which is based on the phenomenon of time dilation that we have just learned about.

Remember what I said about the impossibility of an inertial observer to detect his motion except by looking “outside”. This means that when Bob rushes away from Alice, not only will she see his “photon clock” slow down due to time dilation, but every periodic activity that happens to Bob must slow down in proportion! This includes, of course, Bob’s biological clock - the one that tells his heart when to beat, his lungs when to breathe, his stomach when to send hunger signals to the brain, ... This way, Bob will not feel this slowing down at all! Otherwise, all he has to do is to time his photon clock with his pulse (something very much like what Galileo did with the hanging wall-lamps in the tower of Pisa), to find out that it is running slow - and hence detect his motion, without looking at Alice or anyone else! the upshot of all this is that when Bob is in high speed motion with respect to Alice, his bodily functions

[2]Actually, there is one more effect that makes the clocks on board of the satellites run differently than the earth based clocks. The fact that the gravitational field of the earth is a lot weaker at the satellite than on the surface has an effect comparable in size to that due to (3.1) - and the GPS system has to compensate for both the effects! The gravitational time dilation is one of the important predictions of Einstein’s general theory of relativity (GTR). Thus, the GPS provides confirmation, on a daily basis, of both of Einstein’s relativity theories!

slow down - and he ages less rapidly!

Now, imagine that Alice and Bob are two twins[3]. On their twentieth birthday, Bob sets out on a high speed intergalactic journey, while Alice stays behind on earth. Fifty years goes by, by Alice's clock, when Bob returns home. Alice is now at the ripe old age of seventy - Bob has a great deal of difficulty recognising her. To Alice's eye, however, Bob has changed very little, - he is very much the young man who had left. The reason, of course, is that Bob's clocks, including his biological one, has been running slowly all this while, so that he has aged only five years during his round trip! So, now when the two twins hug each other, one is seventy years old, while the other is only twentyfive!

The scenario above may seem absurd to us - but is it paradoxical? No! The only reason we have any difficulty in accepting this scenario is because our small-speed experiences condition us to the notion of time being the same for everyone. We don't see anything like this happening around us, but today we have ample experimental evidence to show that this is exactly what happens!

So, where is the paradox? Just look at the thing from Bob's point of view. He sees Alice recede from him and then come back, at the same very large speed! Since the laws of physics are the same for him as well, he concludes that it is Alice's clock that must run slow, and thus if he has aged five years during the trip, Alice must have aged by six months only! So, instead of being twenty five and seventy at the end of the trip, Bob and Alice should be twentyfive and twenty and a half only, respectively! This is the twin paradox - where the final outcome of which twin is older than the other one appears to depend on which of the two friends is

[3]I could have invented a new pair to be the twins in our paradox - but I don't want to clutter up the story with too many characters!

thinking about it!

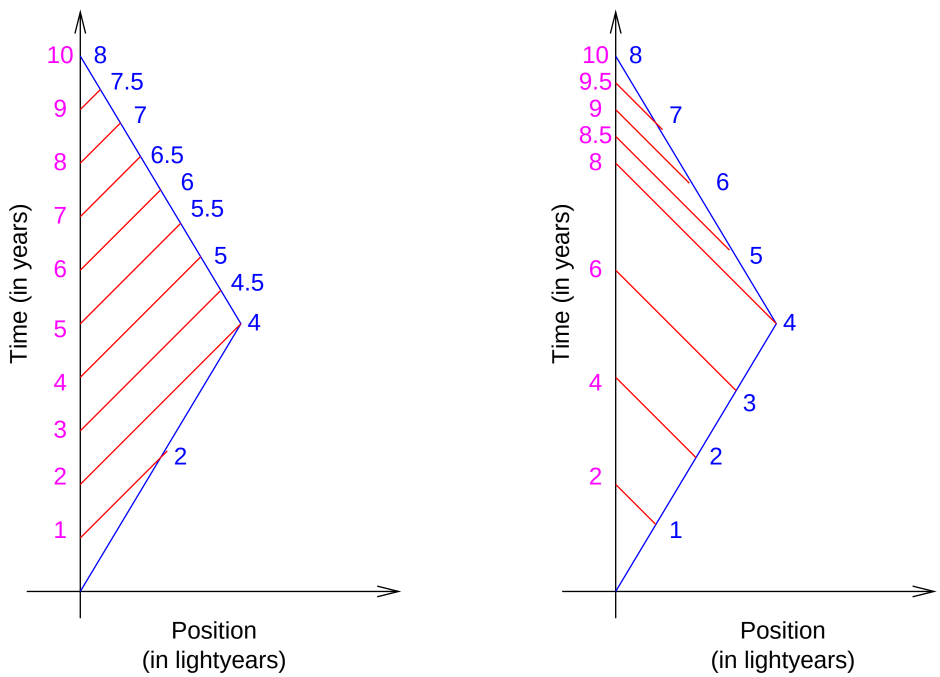
This paradox - one that has bothered a lot of stalwarts in relativity theory - has a very simple resolution - there is just no paradox! The whole issue hinges on the fact that that Bob has just as much a right to draw conclusion on Alice's clock as she has on his. Why? Because the basic principle of relativity says so! Well ... not quite - the basic principle of relativity is that all inertial observers are equal - not that all observers are equal! Alice, the stay at home twin, is inertial - but not Bob. Being inertial, the conclusion that Alice draws is correct. So, Bob's conclusion about Alice being younger than him is simply - wrong!

So, the whole resolution of the twin paradox rests on the fact that Bob accelerates during his journey with respect to Alice, once each at the beginning and the end of his journey, and once at the midpoint in order to turn back. Indeed, I can get rid of the initial and final acceleration from the story by getting Bob to pass by Alice on his rocket (neither start from or stop at her location on earth), but the acceleration at the middle is a must. Note that this acceleration is over a very brief duration compared to Bob's entire journey (and indeed it can be made as brief as you like by making the rockets stronger and stronger) - but the entire crux of the time difference lies in that brief instant.

One thing that has worried a lot of people about this is that for most of his journey Bob is just as inertial as Alice - so they find it difficult to figure out how their conclusions about the time it takes differ so much. However, Hermann Bondi has come up with a great analogy to show why this is not something one should bother about. If Alice had driven straight from a city **A** to a city **B**, while Bob makes the trip **A** to **C** and then on to **B**, then, when they meet Bob's odometer reads a lot more than Alice's. This despite the fact that during most of his journey, Bob had driven just as straight as Alice did! The one brief period at **C** when Bob makes a turn is ultimately responsible for the length of Bob's path being a lot longer than Alice's! Similarly, the one brief period of acceleration during Bob's intergalactic round trip is what makes Alice's clock read a lot more than Bob's. The only reason why people never bat an eyelid in the former case, while they are all full of wonder at the latter is that our intuitions, which are honed by what we see around us in a world of slowly moving objects, are not prepared to grasp the concept that time can be different for different people!

One argument that many people come up with at this stage is that from Bob's point of view it is Alice that accelerates, so it is equally correct to say that Bob is inertial and Alice is not - and hence conclude that Bob is correct about their respective ages and not Alice. This argument falls flat simply because the issue of which observer is inertial and which is not is not is not a relative matter! Remember, Newton's first law asserts that their exists at least one inertial observer for whom all force free objects are also acceleration free. Anyone moving uniformly with respect to this inertial observer is inertial, anyone who accelerates with respect to it is non-inertial! Moreover, although an observer can not make out whether she is moving or not without looking out when in uniform motion, she can easily figure out whether she is accelerating with respect to an inertial observer. In the brief moment of reversal during Bob's journey, Bob's accelerometer must have registered a large reading, while Alice felt nothing at all! So, Alice is the one who is inertial - and their is no ambiguity about this whatsoever!

Although we have disposed off the paradox simply by denying its existence, it is natural to feel a bit uncomfortable about this. It is true that the clocks of both friends run slow compared to the other one for most of the time, so it may be somewhat disconcerting that the final outcome is so asymmetrical. In order to make this a bit easier to digest, let me show you another way of understanding what's happening during the trip.



(a) Alice's signals to Bob

(a) Bob's signals to Alice

Figure 3.2: The twin paradox

For the sake of convenience, I will assume that Bob does not move as fast as in our example above, but with only a speed of $0.6\,c$. This gives us a time dilation factor γ of 1.25, so that a journey that takes only 8 years according to Bob's while Bob's clock takes 10 years according to Alice. Before Bob sets out on his journey, the two friends decide on a strategy by which they can keep track of each other's ages. They both agree that on their birthdays each year they will send out a light flash towards the other one. So, all they have to do is keep count of the flashes they receive, to know how old the other one is. These flashes are shown in the spacetime diagrams of figure 3.2.

As you must have figured out, the Doppler factor will play a major role in helping us keep track of when each friend receives the others signals. With a value of $\beta = 0.6$, Bob's K factor with respect to Alice for the outward journey is 2, and that for the inward journey is $\frac{1}{2}$. Of course, the respective factors for Alice with respect to Bob are $\frac{1}{2}$ and 2, respectively.

Let us consider the flashes that Alice sends out first. Bob's Doppler factor of 2 ensures that her annual signals reach Bob at gaps of two years according to his clock. So, during the outward journey which lasts for four years according to Bob, Bob receives only two of Alice's signals. At the moment he receives' Alice' second signal, Bob fires his rockets and turns back home. Now his Doppler factor is $\frac{1}{2}$, so that Alice's next signals reach him at gaps of six months. For the next four years of his journey, then, Bob receives eight signals from Alice, the final one getting to him just as he is entering the earth again.So, Alice sends out a total of ten signals in all - Bob receives all ten, and they both agree that Alice has aged by 10 years during the round trip.

Let's take a look now at the signals that Bob sends Alice. As long as Bob is moving away from Alice to the right, the Doppler factor of Alice with respect to Bob is $\frac{1}{2}$. Remember, though that Bob's light signals sent to Alice are moving to the left, and this means that Alice will see them at a gap that is twice as big than the one that Bob sends them out at. Thus, Bob sends out a total of four signals during his outward journey, and the fourth of these reaches Alice after eight years (according to her clock). Once Bob turns back, Alice's K factor with respect to Bob becomes 2 - but this means that she is going to receive Bob's yearly signals at gaps of six months!Bob sends out four signals while coming back. Alice receives all of them crammed up in the last two years of her ten. Thus, Bob sends out a total of eight signals, Alice receives all eight! Everything checks out, and they both agree that Alice is ten years older than when the journey started, while Bob has aged by only eight!

3.1.6 The writing on the clocks

I have been careful so far to ensure that our observer use only one clock, the one that he or she carries, to time various events. This way, of course, each observer can directly time only the events that occur at his or her origin. For all other events they have to rely on the radar rule, (2.2-a). This , of course, is not the only approach that one could take to time measurements - many others are possible. One way, that Einstein himself was fond of using is to assume that each observer carries with him or her an infinite number of fixed clocks, one at each point in space. Thus, if Alice was equipped with such an infinite number of clocks, all she would have to do to locate a particular event in spacetime is to look up the reading in her clock that is next to where the event occurs for the time, while the fixed coordinates of that particular clock tells her where that particular event occurs.

If such a way of measuring space and time is adopted, we will have to change our way of defining proper time measurement slightly. In this new way of doing things, proper time is, once again, any time measurement done using a single clock, while any other observer who must use two clocks to measure the time interval between two events (because the two events occur at different places with respect to him) does not measure the proper time interval.

Time dilation leads to a puzzle similar to, but not quite the same as, the twin paradox for two observers equipped with this infinite set of personal clocks. Just imagine that Bob, who has his own private array of fixed clocks, is rushing past Alice and her clocks at the speed of $0.6\,c$. When the two friends are just abreast, their clocks both read 0. After a while Bob glances at his clock and finds that it is reading 8 seconds. He also notices that Alice's clock right next to him is now reading 10 seconds. To Bob, the interval that has elapsed is eight seconds long, whereas to Alice its length is 10 seconds. Of course, Bob uses the same clock for both the readings of 8 seconds and 0 seconds - so his reading is the proper time duration of this interval. Alice, on the other and must use two of her clocks to time this interval - the first one at the origin, and the second at a distance of $0.6\,c \times 10$ s = 6 light-seconds away from her. You should check that the two friends' statements about the length of the time interval is consistent with our time dilation equation, (3.1).

So, where is the puzzle? Note that both friends must agree that at the first event (Bob crossing Alice) their clocks were reading 0 s. They must also both agree that when Bob next looked at his clock (the second event), their adjacent clocks were reading 10 s and 8 s, respectively. All this fits our results beautifully. Hang on! Just try to figure things out from Bob's point of view. To him, it is Alice and her clocks that are moving away, with a speed of $0.6\,c$ to the left. Hence, he should be seeing Alice's clocks running slow, by the same γ factor of 1.25! So, shouldn't Alice's clock be showing 8 s $\times \frac{1}{1.25} = 6.4$ s, when his clock is showing 8 s?

In other words, the puzzle is that while relativity insists that both of the friends will see the other one's clocks running slow, the writing on the clocks seem to insist that it is Bob's clocks that have slowed down - not Alice's! Note that in this case both friends are equally inertial - so that we do not have the escape route that we used for the twin paradox. To see how to resolve this issue you will have to wait a while until section 3.3 - where we will meet a consequence of relativity that runs even more counter to our intuition than time dilation!

3.1.7 The case of the immortal muon

In case the paradoxes in the last couple of sections have shaken your resolve, cheer up! For I will be talking in this section of arguably the most dramatic confirmation of time dilation. This is provide by the mystery of the undying muon. No, I am not referring to some supernatural mystery, but to a natural phenomenon that physicists see quite regularly in their laboratories. Without time dilation to help us out, though, this would have been just as difficult to explain as the ghost stories that we don't believe in (I hope) but definitely enjoy!

The muon is an elementary particle - one of the many subnuclear constituents of matter. One of the places where the muon is produced is in the upper atmosphere, about 5-6 kms above the surface in collision processes involving cosmic rays. These muons reach the surface on a regular basis - to be detected by particle detectors that we have lined up for the purpose.

By itself, this would not have been a big surprise. What makes the arrival of the muons a mystery is that they are unstable, decaying into other particles. How long do they live? What helps us here is that muons can be produced in the laboratory by nuclear processes. The half life of such lab bred muons can be measured, and turns out to be about 2×10^{-6} s. In this small a time, even light can travel only 600 meters! How, then, can the muons produced five-six kilometers above the surface survive till they reach the earths surface?

The answer lies, of course, in the fact that the muon decays by its own clock, one that is slowed down by quite a big γ factor compared to the ones that we carry. Put in a more technical language, in the inertial frame in which a muon is at rest, it does decay with the half life of 2×10^{-6} s. This is the so called **proper half-life** of the muon. However, the muons produced in the upper atmosphere are travelling so fast (at nearly the speed of light) relative to us that their clock runs slow by a factor of around $\gamma \sim 10$, allowing most of them them to get to the surface undecayed.

So that's it - if the muon can make it to the ground from 5-6 kms above, it is all because of the fact that the clock that tells it to decay runs slowly compared to our earthbound clocks. Before you start to celebrate the solution of this case of the undying muon, though, let me just show you the whole life of the muon from the muon's point of view - or rather, from an inertial frame in which the muon is at rest. The muon, in its own frame lives for 2×10^{-6} s. From the muon's point of view, it is the earth that rushes at it. Does the earth move 5-6 kms or more in this small a time? If it did, it is obvious that the earth would have to move at many times the speed of light!

So even though the fact that the cosmic ray muons reach the surface can be explained using time dilation from the earthbound observers reference frame, it still remains a mystery from the muon's frame. The only way that you can solve this is if what we earthlings measure as 5-6 kms, is only a few hundred meters long for the muon! This brings us to the topic of the next section - the fact that moving rods get shortened!

This illustrates a very important theme in relativity theory - two different observers may differ in their explanations about physical events, but will have to agree on what happens! That the muon reaches the ground is something both the earth based observer and the fellow riding in with the muon will have to agree about (although the latter will state it as the ground reaching the muon!). However, though the earth based observer will attribute the muon's amazing longevity to time dilation, the muon rider will explain the facts by invoking the concept of length contraction.

3.2 Length contraction

The case of the muon that refuses to die, suggests, when you are trying to explain this from the muon's point of view, that there must be something going on with moving lengths. Indeed, this is a problem that you may have noticed even earlier. In subsection 3.1.6 we saw that after Bob has travelled for 10 seconds with respect to Alice at a speed of $0.6\,c$, their respective clocks read 8 s and 10 s, respectively. How far apart are they from each other? As we have already noted, Alice reckons that Bob is $0.6\,c \times 10$ s = 6 light-seconds away from her. What about Bob? To him, Alice is moving away with the same speed in the other direction, but she has only been moving for 8 s since crossing him! Thus, Bob must insist that at this moment she is $0.6\,c \times 8$ s = 4.8 light-seconds away only! How can the two friends disagree about how far they are from each other at the same instant? Once again, we see the hint that like time measurements, in STR there is something subtle about space measurements, too!

3.2.1 A moving rod gets shortened

At the end of the last section we have seen that moving rods must contract if we are to apply physics equally in all inertial frames.To see that this really occurs, let us go back to the radar method. This time, as Bob rushes away from Alice, he carries a rod with him, directed along the relative motion. The length of the rod, according to Bob, is l_0. Anticipating the fact that the length may be different to other observers, we call the length in Bob's frame (where the rod is at rest) the **proper length** of the rod. The question that I am going to address now, just how long will Alice claim this rod to be?

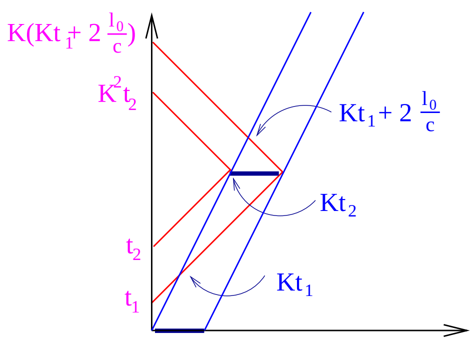


Figure 3.3: Length contraction

To measure the length of the rod Alice must determine the position of both ends of the rod simultaneously. Let me stress that this demand of simultaneity has nothing to do with Einstein, even good old classical physics will tell you that you can't claim the difference between the coordinates of two ends of a moving object to be its length - unless the two measurements are made at the same instant, the object's motion will make your length measurement go all wrong. The only case where you don't have to bother to read the position of both ends together is when the object is not moving with respect to you. For Bob, then, it does not matter when the two ends are measured, he will always find that the difference in coordinates is l_0!

To carry out her measurement, Alice sends out two beams from her flashlight, as shown in figure 3.3. The first one, send out at time t_1 bounces off the far end of Bob's rod and returns to her. the second flash, send out at time t_1 according to Alice's clock, bounces off the near end of the rod (which is at Bob's location) at a time Kt_2 (as read by Bob's clock) to reach Alice at a time K^2t_2 (according to her clock).

We have to exercise a bit more careful to figure out the timings on the first beam. It reaches Bob, of course at a time Kt_1 according to Bob's clock. Then it moves on to bounce off the other end of the rod. Since the speed of the beam is c with respect to Bob, too, the time it takes to come back to him must be $\frac{2l_0}{c}$. Thus, on its way back, the beam meets Bob at the time $Kt_1 + \frac{2l_0}{c}$, once again, as read by Bob's clock. This beam reaches Alice at a time (read by her) that is larger by the factor K again, at $K\left(Kt_1 + \frac{2l_0}{c}\right)$. Figure 3.3 shows the times that Bob measures in blue, while those that Alice measures are shown in purple.

The groundwork is now done - all that is left is to use our basic equations (2.2-a) and (2.2-b) to figure out the difference in time and position of the two bounces as seen by Alice. It is easy to see that the time gap between the two bounces is

$$\begin{aligned}\Delta t &= \frac{1}{2}\left[K\left(Kt_1 + \frac{2l_0}{c}\right) + t_1\right] - \frac{1}{2}\left(K^2+1\right)t_2 \\ &= \frac{1}{2}\left(K^2+1\right)(t_1 - t_2) + \frac{Kl_0}{c} \qquad (3.6)\end{aligned}$$

while the gap in position is

$$\begin{aligned}\Delta x &= \frac{c}{2}\left[K\left(Kt_1 + \frac{2l_0}{c}\right) - t_1\right] - \frac{c}{2}\left(K^2-1\right)t_2 \\ &= \frac{c}{2}\left(K^2-1\right)(t_1 - t_2) + Kl_0 \qquad (3.7)\end{aligned}$$

Now, Δx above will be the length of the rod as measured by Alice, l, only if the two bounces occurred simultaneously, according to her. Thus, we must have $\Delta t = 0$. According to (3.6) this means that

$$t_1 - t_2 = -\frac{2K}{K^2+1}\frac{l_0}{c} \tag{3.8}$$

This means that l is given by

$$l = Kl_0 - Kl_0\frac{K^2-1}{K^2+1} = \frac{2K}{K^2+1}l_0$$

The factor $\frac{K^2+1}{2K}$ is one we have met before, in (3.2). So, we can rewrite the formula for l as

$$l = \frac{1}{\gamma}l_0 = l_0\sqrt{1-\beta^2} \tag{3.9}$$

A word about the name of this effect. Einstein's theory shows us that this contraction of the length of a moving object is a direct consequence of the way it forces us to rethink issues of measuring length and time. However, before Einstein, Fitzgerald and Lorentz had come up with the notion that moving objects contract in the direction of motion, essentially so that they could explain the null effect of the Michelson-Morley experiment. Their prediction was more of a mathematical device that was based on the ether theory than what we think of length contraction about today, but in their honor this relativistic effect is called the Lorentz-Fitzgerald length contraction.

The same factor comes into play in the length contraction of moving rods as in time dilation. Think about it a bit and you will realize that that's exactly what we need to resolve the difference between the two friends about their distances that we talked about at the beginning of this section. At a time when Bob is 6 light-seconds away from Alice, according to Alice - Alice is only 4.8 light-seconds away according to Bob. There is nothing wrong with this - after all, Alice reckons Bob's distance by noting the coordinates of her clock that is just next to Bob at that instant - this space interval is fixed as far as Alice is concerned, but it is moving from Bob's point of view. So, Bob sees it shortened by a factor of $\frac{1}{1.25} = 0.8$ - which is why his reckoning of the distance is 4.8 m!

Length contraction, again, is just what the doctor ordered for the case of the undying muon! From the muon's point of view, it decays in its proper lifetime. In which the ground rushes towards it to cover the distance from the upper atmosphere to the earth's surface. This is a distance that we on the earth measure to be about 5-6 kms, but to the muon it is shortened to about 500-600 meters - just the sort of distance the earth is expected to cover before the muon decays!

Of course, Alice does not *have* to measure the two endpoints of Bob's rod to get its length correct. If there is a time gap of Δt between the two measurements, all she has to do is correct for the distance $v\Delta t$ that the rod would have moved in the interim. Thus, another way to figure out the length l would be

$$l = \Delta x - v\Delta t$$

I leave it as an exercise to you to show that if you use (3.6) and (3.7) in this equation, t_1 and t_2 will cancel out, leaving exactly (3.9) behind.

By now you should have learned enough relativity to expect that just as Alice says that the rod carried by Bob is shortened, so would Bob say that any rod that is fixed with respect to Alice is shorter than what Alice measures it to be, and shortened by exactly the same factor to boot!

Note that in the last statement I was being very, very careful. I said 'says' - not 'sees'. To see an object, light must travel from that object to ones eyes. What you see is the result of light falling on your retina at a given instant. Light travelling from different parts of an object actually travels for different lengths of time before reaching your eye - so they must have started out at different times. Of course, this does not bother us in our daily lives - the times involved are just to small to make a difference. However, for rapidly moving objects, this small time difference could result in large shifts - leaving quite a different image of the object! The visual appearance of rapidly moving objects is quite an involved topic - even the great George Gamow got it wrong in the first edition of his wonderful "Mr. Tompkins in wonderland"! I will say a little bit on this topic a while later.

3.2.2 What about the other directions?

So far we have been sticking to one spatial dimension. This is as good a point as any in which to start worrying about the other directions as well. We have seen that a moving rod gets shortened, but that proof was only for a rod that is aligned along the direction of motion. Indeed, the more perceptive among you may be beginning to feel that there was something wrong with my photon clock derivation of time dilation. what I had assumed there is that the vertical distance between the two mirrors in figure 3.1 stays the same for both Alice and bob at l_0. At that stage, we had not heard of length contraction, so that we had no reason to be worried about this assumption! Now that we have learned that moving objects contract, it is only natural to be quite suspicious of this assumption.

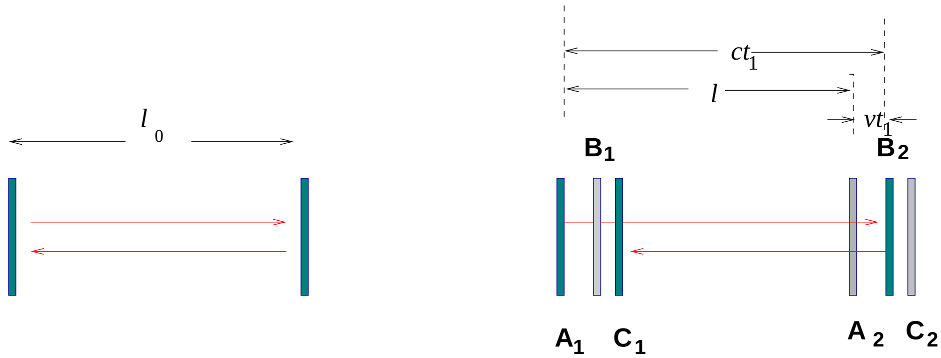
Remember, though, that before talking about the photon clock I had already proved the time dilation equation (3.1) by using the K

calculus. So, there is no reason to doubt the validity of this equation. Indeed, we may even turn the argument with the photon clock over on its head and use it to *prove* that the distance between the two mirrors must remain l_0 for Alice as well as Bob - from the fact that this is the only way we will get the time dilation equation right for this clock! Thus, instead of providing an alternative derivation of the time dilation equation, the photon clock can be regarded as a way of proving the fact that lengths perpendicular to the direction of relative motion of two observers remain unchanged.

Thus, relative motion between two observers shorten lengths in the direction of motion but leave transverse lengths the same. We can leave the matter here - but I can not resist showing you another delightful little proof of the invariance of lengths perpendicular to the motion. Just imagine that there is a vertical wall standing next to the path that Bob takes as he rushes away from Alice. Both Alice and Bob have brushes dipped in color, Alice's her favourite pink, while Bob's is blue. Each friend draws a line at a height of 1 m above the ground on the wall. The question now is - whose mark will be higher up? Now, the fact that there is no difference between left and right is enough to ensure that neither Bob's nor Alice's marks can be higher up than the other's! This leaves us with the only conclusion possible - both friends measure the same vertical distance!

3.2.3 Length contraction via the photon clock

In the last subsection I showed you that the photon clock could be used as a means of proving that lengths perpendicular to the motion are unaffected by it. There, of course, we had placed the clock in such a way that the line joining the mirrors is perpendicular to the motion. It is natural to ask now - "what will happen if we turn



(a) Bob's view (a) Alice's view

Figure 3.4: Length contraction from the photon clock

the clock around so that this line is now parallel to the motion?" Of course, if we use this setup with proper care, this should give us an alternative derivation of the length contraction equation, (3.9).

As far as Bob is concerned, the clock is stationary - so all that light does is travel a distance of l_0 on both halves of the journey, taking a time of $\frac{2l_0}{c}$. What does Alice see? Figure 3.4b shows the whole process of light travelling from one mirror to the other one and back from Alice's point of view. I have indicated the initial position of the two mirrors, when the light beam starts out from the mirror on the left, as A_1 and A_2, respectively. By the time light reaches the right hand mirror, the two mirrors have shifted to B_1 and B_2, respectively. Finally, when the light beam returns to the left hand mirror, their respective positions areC_1 and C_2. The gaps A_1A_2, B_1B_2 and C_1C_2 are each equal, of course, to the length of the photon clock as measured by Alice. Prompted by twenty-twenty hindsight, I will call this length l and not assume beforehand that it is the same length l_0 that Bob sees.

The distances travelled by light and the mirrors are indicated in figure 3.4b, where t_1 is the time, according to Alice, that light takes to travel from the left hand mirror at A_1 to reach the right-hand one at B_2. From the figure it is obvious that

$$ct_1 = l + vt_1$$

which leads immediately to

$$t_1 = \frac{l}{c - v}.$$

In the same way, it is obvious that the time t_2 that the light beam takes to get back to the left hand mirror, now at C_1 is given by

$$t_2 = \frac{l}{c + v}.$$

So, Alice reckons that the round trip takes light a total time of

$$t = \frac{l}{c - v} + \frac{l}{c + v} = \frac{2l}{c} \times \frac{1}{1 - \frac{v^2}{c^2}} = \gamma^2 \frac{2l}{c}. \tag{3.10}$$

So, now we know that Bob reckons that the time taken for the round trip by light is $\frac{2l_0}{c}$, while Alice reckons that it is $\gamma^2 \frac{2l}{c}$. Should these two times be equal? Although in pre-relativity days the answer would have been an unequivocal "Yes!", today we know better! After all, as far as Bob is concerned, the round trip taken by light begins and ends at the same place, namely his origin. So, the time that Bob sees is the proper time interval that elapsed during the trip. So, Alice must see a dilated time, given by $\gamma \times \frac{2l_0}{c}$ - but we already know that this is $\gamma^2 \frac{2l}{c}$. This leads immediately to

$$l = l_0 \times \frac{1}{\gamma} = l_0 \sqrt{1 - \beta^2}$$

- the length contraction formula!

Although we have already achieved our target, it is instructive to look back at the way in which we calculated Alice's time measurements above. If this would have been pre-relativity physics - Alice could have calculated t_1 in this way : "I see light travel to the right at a speed of c, so Bob sees it travel at a rate of $c - v$. The mirrors are fixed as far as Bob is concerned, so to Bob. light just covers the distance between the mirrors. Since Bob sees light cover the distance l (the length of the mirror being the same as the one I measure for him), it must take light a time of $\frac{l}{c-v}$!" Wrong reasoning, as we know today - but the right answer all the same!

Why is the above reasoning wrong? Well, as we all know, Bob does not see light travel at $c-v$, he sees it travel at c! Moreover, the distance between the mirrors that Bob sees is *not* l - but a larger length (the proper length)! Perhaps most importantly, even if this had given me the time that Bob sees the journey from one mirror to another take (it does not!), the result would not have been the same as the time that Alice sees!

Now that we understand why the reasoning above is *wrong*, let us turn to the question -"why, then, is the answer right?" The answer to this question may surprise some of you - it is right simply because it has no other option! After all, why did Alice claim in her pre-STR days that Bob would see the speed of light to be $c - v$? Answer : so that the correct calculation of time (which gave $\frac{l}{c-v}$) matches for both the friends! Put another way, at a time

t Alice sees the lightbeam travel to a point ct away from her, while Bob is vt away from her at that instant. Alice reckons that the distance between Bob and the light spot is $ct - vt = (c - v)\,t$. Dividing this by t gives Alice a speed $c - v$, which is what Alice would have considered the speed of the light spot relative to bob, had she not known that Bob's length and time measurements do not match hers.

This raises another question - exactly why does Bob see the light spot recede from him at speed c? I will give you the answer to this one a bit later, in section 3.4.

3.2.4 The length contraction paradox

If the heading of this subsection sounds too formal to you, feel free to substitute the name under which it is better known - the "car and garage paradox".

Alice has bought a new car that is 10 m long. Her trouble is, her garage, which accommodated her earlier, smaller car is only 8 m long. Having studied the relativistic length contraction, Alice hits upon a nice solution. She asks her friend Bob to drive the car into the garage at a speed of $0.6\,c$. At this speed, the length contraction factor γ^{-1} will be $\sqrt{1-(0.6)^2} = 0.8$ - precisely enough to make the car 8 m long according to Alice! What Alice can now do is slam the door of the garage shut at the same instant that the front end of the car hits the back wall - and so the 10 m car fits into the 8 m garage (I am staying deliberately silent about what happens after that)!

This scenario may seem completely absurd - but that is just because the slow speeds of our daily lives do not prepare us for such things that become apparent only for very fast objects! That is not where the paradox is. The paradox becomes apparent when you look at this from Bob's point of view. To Bob, the car is at its rest length of 10 m, it is the garage that is now moving at $0.6\,c$, making it even shorter - 6.4 m to be precise! So, how on earth can the car fit into the garage, when you see this from Bob's point of view?

To see how this paradox can be resolved, let's examine just what it means when Alice says that the car fits in her garage. Of course, she means that the front end of her car hits the back wall of the garage and the back end crosses the front gate *simultaneously*. There's our clue! As I have been stressing over and over again, the major new thing in STR is the fact that time measurements differ from observer to observer. So Bob does agree with Alice that the front end of the car hits the back wall and the rear end crosses the front gate - its only that he refuses to accept that the two events occur at the same time! You will have to wait until the next section, though, to see that this disagreement between which events occur together and which do not is just enough to explain the difference in the two friend's points of view.

3.3 The relativity of Simultaneity

The fact that amazes people most about STR, apart from things like the twin paradox is perhaps the fact that two events that appear to act at the same time (but at different places) to Alice will not appear to be at the same time to Bob (Remember my story about the waiter and the diners in section 2.1?)! Of course our very first puzzle of me and my friend on the bike should have primed you towards accepting the fact that times may differ from observer to observer - so this may not come as that big a shock to you, after all.

Just how big is this difference in time measurements? You will hardly have to work at all to find this out, since I did almost all the work in the last section! Remember, Alice had to measure the two ends of Bob's rod simultaneously in order to measure its length l. Do cast your mind back to how Alice measures the length - also, refer to figure 3.3. According to Bob, the two bounces from the two ends of the rod happened at times of $Kt_1 + \frac{l_0}{c}$ and Kt_2, respectively. So, the time difference between these two bounces that bob measures is $K\left(t_1 - t_2\right) + \frac{l_0}{c}$. Now, as we have seen in (3.8), the two bounces will be simultaneous to Alice if $t_1 - t_2 = -\frac{2K}{K^2+1}\frac{l_0}{c}$. In this case, the time interval that Bob will see between them becomes

$$\frac{l_0}{c} - \frac{2K^2}{K^2+1}\frac{l_0}{c} = -\frac{K^2-1}{K^2+1}\frac{l_0}{c} = -\beta\frac{l_0}{c}$$

which is a measure of just how relative (*i.e.* observer dependent) the concept of simultaneity is. The minus sign shows that to Bob, the bounce at his origin occurred earlier than the other one. Before we wrap the formula up, let me point out that the variables that are in use there are slightly jumbled up - in particular note that l_0 is the spatial distance between the two bounces as seen by Bob, while it is Alice who sees Bob's velocity to be β (since the result involves the first power of the velocity, it does matter whether we are talking of Bob's speed with respect to Alice, or Alice's speed with respect to Bob!) and more importantly, it is Alice who sees the two bounces to be simultaneous! So, perhaps it would be better to use l, the distance between the events as measured by Alice, instead of l_0 to write the time difference as measured by Bob in the form

$$\Delta t_{\text{Bob}} = -\frac{\beta}{\gamma}\frac{l}{c} = \frac{-\frac{v}{c^2}l}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.11}$$

To wrap up, if Alice sees two events occur at a distance l to be simultaneous, then Bob, who moves at a speed v with respect to her to the right will see a gap of $\frac{\beta}{\gamma}\frac{l}{c}$ between the two events, with the one that is to the left occurring earlier. It is easy to see that had Bob moved to the left, the event that occurs more to the right would appear earlier to Bob.

3.3.1 Einstein's train

Like in the case of time dilation and length contraction, being able to prove equation (3.11) in at least one more way will add to our understanding of exactly what is going on. Let me now turn to a thought experiment aimed at this which comes straight from the horse's mouth - Einstein's famous relativistic train example. Let us imagine that Alice is standing on the side of a railway track, while Bob is rushing by her in a very, very long train that is moving very, very fast. Just at the moment when Bob is abreast with Alice, two flashes of lightning strike at equal distances to their left and right - leaving marks on both the track and the train carriage. Hold on - what does "just at the moment" mean? Since we know that events that are simultaneous to one are not simultaneous to others, such statements must be qualified with the identity of the observer who is seeing the lightning strikes together. Since lengths change for moving observers, too, you may doubt whether the statement "at equal distances to their left and right" needs to be qualified too. Well, it is true that Bob will see a different distance between the marks of the lightning strike and himself, than Alice does - that's length contraction for you! However, as Alice sees the strikes to be at equal distances from her on both sides, then so should Bob - length contraction does not distinguish between left and right. Let me assume that these two lightning strikes are simultaneous with respect to Alice. Then, since Alice is sitting midway between the two flashes, light from the two of them, travelling with equal speeds c, will reach her at the same instant. Of course, Bob, who has been moving all the time to the right, meets the flash coming in from the right hand lightning strike before it reaches Alice, while light from the other strike does not reach him until after it has reached Alice and crossed her. Before the days of Einstein, both friends would have passed this over by saying that light from the right hand strike is moving in a a speed of $c + v$ with respect to Bob, while that from the left hand strike is moving at a speed of only $c - v$, so it is no wonder that even though the two flashes occur together (remember - no one had an inkling that simultaneity is relative before Einstein), Bob sees the right hand strike earlier. However, once we know that light travels at the same speed for all, this escape route is out. Since both flashes are equally distant from Bob, and he sees the right hand flash first, he must reckon that it had occurred earlier! So, the flashes - which are simultaneous with respect to Alice are not simultaneous to Bob - simultaneity is relative!

What is the gap that Bob sees in between the two strikes? Well, it is pretty easy to figure out that Alice will see Bob meet the light from the right hand and left hand flashes after times $\frac{l}{2(c+v)}$ and $\frac{l}{2(c-v)}$, respectively, where $\frac{l}{2}$is the distance of each flash from her. So, Alice will see the time difference between the two lightbeams striking Bob as

$$\frac{l}{2(c-v)} - \frac{l}{2(c+v)} = \frac{vl}{c^2-v^2} = \frac{\beta}{1-\beta^2}\,\frac{l}{c} = \gamma^2\beta\,\frac{l}{c}$$

What, then, is the time difference according to Bob? Note that Bob is present at both the events - so that the time that he measures is the proper time interval between the two events. This means that Bob's value for the time interval is a smaller by a factor of γ. Hence Bob sees the right hand flash $\gamma\beta\frac{l}{c}$ ahead of the left hand flash. Since he is at the midpoint of the position of the thunder-strikes, he will claim that the right hand flash occurs earlier, by a gap of $\gamma\beta\frac{l}{c}$. This, of course is the result that I had shown you a while ago using the radar method.

Using an argument like this we can cover the case that's not so easy to deal with in terms of the radar method. What if the train was moving not along the line joining the two points where lightning had struck - but rather in a direction perpendicular to this line? It is rather easy to see that the two flashes would have reached Bob at the same time in this case, since he keeps on staying equally distant from their sources. Thus Bob and Alice will both agree that the flashes are simultaneous!

What if Bob had moved off at an angle to the line joining the flashes? It seems reasonable to expect that in this case, again, Bob will see the flash towards which his velocity is slanted to occur earlier - you might even be so bold as to suggest that for this case, too, the equation (3.11) will continue to hold, except that instead of the the product vl - you must have the scalar product $\vec{v}\cdot\vec{l}$ - so that the equation becomes

$$\Delta t_{\text{Bob}} = -\gamma\frac{\vec{v}\cdot\vec{l}}{c^2} \tag{3.12}$$

It is easy to check that this general equation definitely matches the results for the two cases that we have considered so far - so that you can have a degree of confidence in it.

You can check whether (3.12) is correct with a little bit of extra effort. Figure ** shows the situation. It is easy to see that Alice will see the flashes from the right hand thunder strike and left hand thunder strike reach Bob at times $t_1$and t_2, respectively, given by

$$\begin{aligned} ct_1 &= \left|\vec{v}t_1 - \frac{\vec{l}}{2}\right| \\ ct_2 &= \left|\vec{v}t_2 + \frac{\vec{l}}{2}\right| \end{aligned}$$

Using $\left|\vec{A}+\vec{B}\right|^2 = A^2 + B^2 + 2\vec{A}\cdot\vec{B}$, you can rewrite the above as

$$\begin{aligned} c^2t_1^2 &= v^2t_1^2 - \left(\vec{v}\cdot\vec{l}\right)t_1 + \frac{l^2}{4} \\ c^2t_2^2 &= v^2t_2^2 + \left(\vec{v}\cdot\vec{l}\right)t_2 + \frac{l^2}{4} \end{aligned}$$

Taking the difference of these two equations immediately gives

$$\left(c^2 - v^2\right)\left(t_1^2 - t_2^2\right) = -\left(\vec{v}\cdot\vec{l}\right)\left(t_1 + t_2\right)$$

Giving

$$t_1 - t_2 = -\frac{\vec{v}\cdot\vec{l}}{c^2 - v^2}$$

- which is the difference in the times at which the two light flashes reach Bob. This doesn't match equation (3.12) - but that's only because it is the time interval as reckoned by Alice. Bob, of course, measures the proper time interval - throwing in the correction factor of $\sqrt{1-\frac{v^2}{c^2}}$ gives you exactly (3.12)!

3.3.2 The writing on the clocks revisited

In subsection 3.1.6 I had introduced to a puzzle that involved the symmetry of time dilation. In brief, if both Alice and Bob sees the other's clock run slow, then how do you explain the fact that Bob sees Alice's clock adjacent to him read 10 seconds when his own shows only 8 seconds?

When Bob's clock showed zero, Alice's clock right next to him had shown zero, too. Thus it might seem to Bob that the time interval that elapsed in between is 10 seconds according to Alice's clocks, whereas his own shows only eight seconds. This of course would mean that Alice's clocks are moving faster than Bob's!

Think about it a bit and you will find that there is a slight gap in the argument. Alice's measurement of the time interval involves two different clocks here - and even in our daily lives (let alone for situations where things move around at nearly the speed of light) we would not subtract the readings on two clocks to get the time interval between two events, unless we were sure that the clocks are set right. In other words, we must be sure that the clocks are *synchronized* - they must both show zero at the same instant. The same instant? This means of course that both had to read zero simultaneously! There are many ways in which this can be done - and since Alice is relying on her clocks to give her accurate time intervals, we take it for granted that she has managed to synchronize her clocks by some means.

Alice, then, does claim that the time interval that elapses between the two events is 10 seconds. For Bob's measurements, of course, we do not have to worry about synchronization - he is using the same clock to time the two events - just subtraction of the readings will do.

Let me now show you what the whole affair looks like to Bob.

He of course can not deny the readings on the clocks, what he will deny though is that the interval as measured by Alice's clocks is really 10 seconds! As far as he is concerned, Alice's clocks are *not* synchronized. You should have expected that - synchronization involves setting the clocks at zero simultaneously, and events that are simultaneous to Alice (who did the synchronization on her clocks) are not simultaneous to Bob!

Now, Alice's two clocks are separated by the distance that Bob travels at a rate of $0.6\,c$ in the 10 seconds that Alice sees as the length of the interval - this of course is 6 light-seconds.Now, think about Alice setting her clock at her origin and her clock 6 lightseconds away. These two events, that occur at the same time according to her, are separated by a time gap given by equation (3.11) as

$$0.6 \times 1.25 \times 6 \text{ s} = 4.5 \text{ s}$$

with the clock at Alice's origin set later. Since the two friends clocks at their respective origins showed zero at the same time, Bob will have no option other than saying that the second clock being used by Alice has been set too early, by 4.5 seconds.

To compare how fast Alice's clocks are running, we must focus on the time gap between two events as read by any one them. Let me take the clock that is 6 light-seconds away,k and consider the two events when it was set to zero - and when Bob came up just abreast to it. Of course, this clock reads a gap of 10 seconds between them. However, to Bob, the first event occurred, not at the time origin, but at $t = -4.5$ seconds, while the second one occurs at $t = 8$ seconds. Hence, as far as Bob is concerned, the gap between the two events that he sees is not 8 seconds at all, but 12.5 seconds. No wonder, then, that he says that Alice's clock is going slow, by a factor of $\frac{12.5}{10} = 1.25$!

3.3.3 The length contraction paradox revisited

Remember Alice's problem of getting her 10 m long car into her 8 m long garage (subsection 3.2.4) ? She could get Bob to drive at the car at $0.6\,c$ to get it to fit in the garage - the trouble was, Bob would see the car to have a length of 10 m, while to him the garage will be shortened to only 6.4 m!

Of course, when Alice says that her car just fits into her garage, she means that the front bumper of her car touches the rear wall, the back of the car just crosses the gate. This means that two events - the front end of the car touching the rear wall of the garage and the back end crossing the gate are simultaneous to Alice. By now you must have caught on to the story. These two events that occur simultaneously at a distance of 8 m to Alice (that, remember, is how long the garage as well as the car is to her) occur at a time gap of $0.6 \times 1.25 \times 8$ m$/c = \frac{6\text{ m}}{c}$ according to Bob. Now, Bob sees himself sitting still in the 10 m long car, with a 6.4 m long garage rushing in at him at a speed of $0.6\,c$. How far does the garage move in the $\frac{6\text{ m}}{c}$ time gap that Bob sees between the two events - exactly 3.6 m - the gap in size between the car and the garage! So, everything falls into place!

Come to think about it, by demanding that the garage has to be 8 mlong in order that the car fits into it is actually too restrictive - a smaller garage will do! After all, all Alice needs to do is shut the door on the back of the car before the information that the front end of the car has crashed into the rear wall gets to her. If the garage has a length L, it will take the news at least $\frac{L}{c}$ amount of time to reach her, in which the car (or rather, the back end, which does not know that the front end has crashed as yet!) gets to travel a further distance of βL. So, the car manages to get into the garage even if its length is $L + \beta L = (1 + \beta)\,L$ according to Alice, that is, even if its proper length is $L_0 = \gamma\,(1+\beta)\,L = \sqrt{\frac{1+\beta}{1-\beta}}L$. Note that the Doppler factor K makes a rather unexpected appearance here, can you figure out why?

So, in order to fit a 10 m long car moving at $0.6\,c$, the garage needs be only 5 m long!On the other hand, if Alice has a 8 m long garage, she does not have to get Bob drive at the breakneck speed of $0.6\,c$, a mere $0.22\,c$ will do!

3.4 Addition of velocities

One thing may have been vexing you again and again - just how does the speed of light turn out to be the same for all observers? Or, why is it if Bob sees Charlie move away from him at a rate of $0.6\,c$, while he is moving away at the same rate from Alice, Charlie's speed as seen by Alice is not $1.2\,c$, but rather, something smaller than c? To see the answer, let me take you back to the radar method once again - you will find that finding out Charlie's speed as seen by Alice is very simple, indeed, if we think in terms of the Doppler factors instead of conventional speeds.

Consider a situation where Alice, Bob and Charlie where together at their respective zero times. Bob is moving away with a speed v_1 and a corresponding Doppler factor K_1 with respect to Alice. Charlie, on the other hand, is moving away with speed v_2 (and Doppler factor K_2) with respect to Bob. The question is, what is the velocity v_3 of Charlie with respect to Alice? Of course, since K and β are related to each other via the equations (2.3) and (2.4), we can just as well ask, what is the value of the Doppler factor of Charlie with respect to Alice, K_3?

By now you know enough relativity to know that the answer cannot be the simple classical result $v_1 + v_2$ - the formula for relativistic addition of velocities must be different! To find out what it is, just imagine that Alice flashes her flashlight at a time τ (according to her Clock) in the direction of Bob and Charlie. Bob receives the flash of light at time $K_1\tau$ (according to his watch), which then proceed to meet Charlie at a time $K_3\tau$ (this time its Charlie's watch). Now, Bob could just as well have claimed that he had sent out the signal at a time $K_1\tau$, which should have reached Charlie when the latter's watch reads $K_2 \times K_1\tau = K_1K_2\tau$. This means that we must have

$$K_3\tau = K_1K_2\tau$$

so that the relativistic addition of velocities simply becomes the multiplication of Doppler factors

$$K_3 = K_1K_2 \tag{3.13}$$

Indeed, as we will see over and over again, this simple result is the reason why the Doppler factor is more convenient to use than the speed in relativistic contexts. We have already seen, near the end of section 2.4 that the Doppler factor is more directly amenable to experimental determination. Be that as it may, speeds being more familiar, we would definitely like to write (3.13) in terms of v_1, v_2 and v_3. Using equation (2.4), this is readily done:

$$K_3^2 = K_1^2K_2^2$$

implies

$$\frac{1+\beta_3}{1-\beta_3} = \frac{1+\beta_1}{1-\beta_1} \times \frac{1+\beta_2}{1-\beta_2}$$

which leads readily to

$$\beta_3 \quad = \quad \frac{(1+\beta_1)(1+\beta_2) - (1-\beta_1)(1-\beta_2)}{(1+\beta_1)(1+\beta_2) + (1-\beta_1)(1-\beta_2)}$$

$$= \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2}$$

or, in terms of the more familiar speeds

$$v_3 = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}} \tag{3.14}$$

This, then, is our relativistic formula for addition of velocities.

Now that we have found out our formula for the addition of velocities in STR, let us check some of the statements I have been making so far. First, let us see what the scenario will be if in place of Charlie we have a beam of light. Then $v_2 = c$ and our formula gives us

$$v_3 = \frac{v_1 + c}{1 + \frac{v_1 c}{c^2}} = c!$$

- thus, the speed of light is c with respect to Alice, too. After all, since I have assumed that the speed of light is the same for all observers as the basis of our radar method, I would have been very surprised indeed if anything else had come out of our calculation! You can easily check that if Bob sees a light beam moving to the left, $v_2 = -c$, then $v_3 = -c$, too. Remember what I had said about being unable to outrun light?

The fact that c remains c under the velocity addition rule becomes even more obvious in terms of the Doppler factors. Remember that light travelling to the right (left) has a Doppler factor K_2 of ∞(0) - and this obviously does not change upon multiplication by K_1.

Another result that follows from our rule of addition of velocities is the fact that if Bob sees Charlie move slower than light, so will Alice, no matter how fast Bob moves with respect to Alice. You can easily check that, for example $v_1 = v_2 = 0.6\,c$ leads to

$$v_3 = \frac{0.6\,c + 0.6\,c}{1 + \frac{0.6\,c \times 0.6\,c}{c^2}} = \frac{1.2}{1.36}c \approx 0.88\,c$$

or even that $v_1 = v_2 = c$ leads to

$$v_3 = c \quad !$$

Proving this result, though, involves a slightly lengthy bit of algebra. This result, though, is almost immediate if you think in terms of the Doppler factor instead of the speeds.

3.4.1 The relative velocity formula

The difference between the relative velocity formula and the velocity addition formula is really one of context. You use the latter if you know Charlie's velocity with respect to Bob and Bob's velocity with respect to Alice and want to find out Charlie's velocity with respect to Alice. You will use the former if given both Charlie and Bob's velocities with respect to Alice, you want to figure out Charlie's velocity with respect to Bob. In the notation that I used above, then, you know K_3 and K_1 and want to find K_2. The result is obvious and leads directly to

$$K_{\text{Charlie w.r.t. Bob}} = K_{\text{Charlie w.r.t. Alice}} / K_{\text{Bob w.r.t. Alice}} \tag{3.15}$$

So, the calculation of relative velocity just boils down to the division of Doppler factors. In terms of the velocities, this formula becomes, instead of the classical $u - v$,

$$u' = \frac{u - v}{1 - \frac{uv}{c^2}} \tag{3.16}$$

- where u' is Charlie's velocity relative to Bob, with u and v being the velocities of Charlie and Bob, respectively, with respect to Alice. Check this!

3.4.2 A deeper look into relative velocities

Now that we have the relative velocity formula safely under our grasp, let m try to show you how to understand this result. To give you a flavour of what I intend to do in this subsection, let me first show you how the familiar $u - v$ result for relative velocity in pre-Einsteinian physics can be understood directly in terms of quantities measured by our two friends.

Imagine that you are back in the good old comfortable Newtonian world, in which Alice sees Bob and Charlie receding from her with speeds v and u, respectively, along the same direction. This means that at a time t (you no longer have to specify by *whose* watch - remember that time is universal in Newtonian physics) they are at respective distances of vt and ut from Alice. Thus at this point of time, Alice measures the distance between them to be $(u - v)\,t$. This is the distance that Bob measures between them, too - length being an invariant in Newtonian physics. Now, the time according to Bob, again, is t - the same as that measured by Alice. This gives Charlie's speed according to Bob as $\frac{(u-v)t}{t} = u - v$ - our familiar relative velocity equation!

Now, let us try to analyse the same situation - but this time taking STR into account. The situation is exactly the same as far as Alice is concerned - she does see the distance between Charlie and Bob as $(u - v)\,t$ at a time t - according to her distance and time measurements. The question, then, of course is - just how far does Bob measure Charlie to be, and what is the time according to him?

Now, Bob will certainly measure the distance to Charlie using a measuring scale that is fixed with respect to him, and so length contraction ensures that there is to be a difference between his and Alice's values for this quantity. If the length of the segment of Bob's measuring rod stretches from him to Charlie is D with respect to him, Alice will see it shortened to $D\sqrt{1-\frac{v^2}{c^2}}$. This of course must match Alice's distance reading, $(u-v)\,t$ (a bit of care must be exercised here to keep the logic straight - the two ends of the segment of Bob's rod are at ut and vt according to Alice, but what ensures that the difference between these two readings is the *length* of this segment as far as she is concerned is the fact that both these readings have been taken, according to her, at the same time, namely t ! This is why I can simply use the length contraction formula here.), and so D must be

$$D = \frac{(u-v)\,t}{\sqrt{1-\frac{v^2}{c^2}}}$$

Now that we have the distance to Charlie according to Bob, the next question is, what is the time, again according to Bob, when Charlie is at this distance? We have seen that Alice's clock right next to Charlie is showing a time t at this instant - but by now we know enough to realize that the time that Bob reckons for this must be different. Since Alice's clock is moving past Bob at a speed of v, it must be running slow by a factor of $\sqrt{1-\frac{v^2}{c^2}}$. Just compensating for this will tell us that Bob reckons the time to be

$$T = \frac{t}{\sqrt{1-\frac{v^2}{c^2}}}$$

and this means that Bob will see Charlie's speed to be

$$u' = \frac{D}{T} = \frac{\frac{(u-v)t}{\sqrt{1-\frac{v^2}{c^2}}}}{\frac{t}{\sqrt{1-\frac{v^2}{c^2}}}} = u - v$$

- the same as the classical result! Something must have gone very wrong here - after all, the classical formula for the addition of velocities just can not be right! Apart from anything else - if instead of Charlie, both Alice and Bob had been observing a flash of light - this formula would have told you that they would have seen different speeds - something that runs smack against relativity's basic postulate!

Some of you must have caught on to the source of the trouble by now - I have simply forgotten to take the disagreement between the two friends on clock synchronization into account. Alice says that Charlie has covered the distance ut in time t - because at $t = 0$ he was right next to her, while at time t he is at a distance of ut from her. Though, the two clocks that tell her the time at Charlie's two locations are different, she is sure that subtracting their readings does give her the time interval elapsed because she has synchronized them so that they read zero a the same time. Bob, on the other hand will say that the second clock was set to zero too early - it has a head-start of $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \frac{v}{c^2} \times (ut)$ over the clock at Alice's origin. This means that the time that has elapsed according to Bob is not really $\frac{t}{\sqrt{1-\frac{v^2}{c^2}}}$, but rather

$$T = \frac{t}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \frac{v}{c^2} \times (ut) = \frac{1-\frac{uv}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}} t$$

Using this value for T immediately leads to our law of relative velocities.

As an added bonus we can easily adapt this argument to understand the situation when Charlie is not moving in the same direction as Bob is, with respect to Alice. In this case, the formula for u' certainly stays valid for the x component of Charlie's velocity (remember - the X component is special, because that is the direction in which Bob is moving relative to Alice.). Note that you have to remember to change the u in the term $\frac{uv}{c^2}$ in the denominator to u_x, too - check out the general formula for the relativity of simultaneity (3.12)! This means that we have

$$u'_x = \frac{u_x - v}{1 - \frac{u_x v}{c^2}} \tag{3.17-a}$$

What about the other two components of Charlie's velocity? You know that both Alice and Bob will measure the same value for lengths perpendicular to the direction of relative motion. Thus along the Y and Z directions both will see the separation between Charlie and Bob as $u_y t$ and $u_z t$, respectively (remember that $v_y = v_z = 0$), when Alice's watches show a time t. Although they both measure the same displacement component, the disagreement on time persists, leading to Charlie's velocity components as measured by Bob turning out to be

$$u'_y = \frac{u_y\sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{u_x v}{c^2}} \tag{3.17-b}$$

$$u'_z = \frac{u_z\sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{u_x v}{c^2}} \tag{3.17-c}$$

As you can see, the velocity components perpendicular to the relative velocity of Alice and Bob transforms in a rather complicated fashion. This, despite the fact that distances in this direction are the same for the two friends. Indeed, it is better to say that the complication in this case arises exactly because of this - there is no Lorentz contraction factor that cancels out the time dilation factor!

3.5 Lorentz transformations

All the ingredients necessary for studying relativistic kinematics are now at hand - you can work out all the results with just the few effects that you have learned about so far (just as in the last section I showed you how to understand the velocity transformation formula with just time dilation, length contraction and the relativity of simultaneity). However, the job will be much, much easier if we add to our repertoire of tools the most general one of them all - the rules about how space time coordinates of an event as measured by one observer differ from those measured by another! We are looking, then, for the successor to the venerable Galilean transformations.

Deriving the Lorentz transformations is a very simple matter if we make use of the radar method. Just imagine Alice sending out a beam from her flashlight at time t_1 - which bounces from some distant point and comes back to her at time t_2. On its way out, this beam had crossed Bob at a time Kt_1, while coming back in, it must have met Bob at time $\frac{t_2}{K}$ (this should be obvious to you by now - if not, just study section 2.4 in some detail). Of course, Bob could just as well say that *he* had send out the beam at time Kt_1 and received it after the bounce at time $\frac{t_2}{K}$. So, from our by now familiar rules (2.2-a) and (2.2-b), the spacetime coordinates at which the bounce occurred are :

For Alice :

$$x = \frac{c}{2}(t_2 - t_1) \quad (3.18\text{-a})$$

$$t = \frac{1}{2}(t_2 + t_1) \quad (3.18\text{-b})$$

For Bob :

$$x' = \frac{c}{2}\left(\frac{t_2}{K} - Kt_1\right) \quad (3.19\text{-a})$$

$$t' = \frac{1}{2}\left(\frac{t_2}{K} + Kt_1\right) \quad (3.19\text{-b})$$

The equations (3.18-a-3.19-b) relate the space-time coordinates of the bounce for Alice and bob in terms of the times t_1 and t_2. All we have to do is eliminate these two times from the equations in order to find the transformations that we are after.

All that is left is the algebra. However, there are some interesting things that we can learn on the way. Firstly, it is obvious that (3.18-a) and (3.18-b) leads to

$$t_2 = t + \frac{x}{c} \quad (3.20\text{-a})$$

$$t_1 = t - \frac{x}{c} \quad (3.20\text{-b})$$

which leads immediately to the relation

$$t_1 t_2 = t^2 - \frac{x^2}{c^2} \quad (3.21)$$

What makes the above equation very interesting indeed is that the corresponding product for Bob is $Kt_1 \times \frac{t_2}{K} = t_1 t_2$ - the same as for Alice. So we immediately get

$$t^2 - \frac{x^2}{c^2} = t'^2 - \frac{x'^2}{c^2} \quad (3.22)$$

an equation that, as we will see, has immense importance in relativistic physics.

Now, for the actual derivation! Substituting (3.20-a) and (3.20-b) in (3.19-a) and (3.19-b) immediately leads to

$$\begin{aligned} x' &= \frac{1}{2}\left(K+\frac{1}{K}\right)x-\frac{c}{2}\left(K-\frac{1}{K}\right)t \\ t' &= -\frac{1}{2c}\left(K-\frac{1}{K}\right)x+\frac{1}{2}\left(K+\frac{1}{K}\right)t \end{aligned}$$

which are the transformations that we are looking for. Of course, the factors $\frac{1}{2}\left(K+K^{-1}\right)$ and $\frac{1}{2}\left(K-K^{-1}\right)$ are old friends - they are none other than γ and $\beta\gamma$, respectively. In terms of these, the equations become

$$x' = \frac{x-vt}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.23-a}$$

$$y' = y \tag{3.23-b}$$

$$z' = z \tag{3.23-c}$$

$$t' = \frac{t-\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.23-d}$$

where I have thrown in (3.23-b) and (3.23-c) which simply say that lengths perpendicular to the direction of relative motion are the same for both Alice and Bob. These, then, are the transformation equations that are going to take the place of the Galilean transformations. They had been found on mathematical grounds by H. Lorentz before Einstein put them on a strong physical footing. This is why these are known as the Lorentz transformation equations - and the passage from one inertial observer to another is called a Lorenz transformation.

Our basic postulate of relativity boils down to the mathematical demand that the form of the equations of physics must stay the same under a Lorentz transformation. In a slightly more convoluted fashion, scientists prefer to put this as “The laws of physics should be Lorentz covariant". Just what does covariance mean - and just why am I stressing on that as opposed to, say, Lorentz invariance are issues that I am going to describe in some detail later on.

For the time being let's see how Alice can find the space time coordinates of an event, given the coordinates that Bob has measured for it. In other words, we want to find (x, y, z, t) from (x', y', z', t'). One way to do this is simply to solve the Lorentz transformation equations (3.23-a-3.19-b) to find (x, y, z, t). I leave the job of carrying out the algebra to you. However, there is a very much simpler way of finding the result - all you have to do is to remember our basic postulate of relativity. The equation that connects Alice's readings to Bob's must have the same form as those that connect Bob's readings to Alice's - except that instead of v we must put in $-v$, which is Alice's velocity with respect to Bob. Thus, the so called *inverse* Lorentz transformations are

$$x = \frac{x' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{3.24-a}$$

$$y = y' \tag{3.24-b}$$

$$z = z' \tag{3.24-c}$$

$$t = \frac{t' + \frac{v}{c^2}x'}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{3.24-d}$$

You should check for yourself that solving the Lorentz transformation equations for (x, y, z, t) yields the same results. Also check that equations (3.23-a) and (3.23-d) directly lead to (3.22).

3.5.1 Everything from L. T.

Let me point out now that a more traditional presentation of the special theory of relativity would have started out by first proving the Lorentz transformation equations and then deriving all the other phenomena that we have discovered from them. By the way, the more traditional presentations would also have called the two observers S and S' - I personally feel that Alice and Bob is much nicer!

Once you have (3.23-a-3.19-b) and (3.24-a-3.24-d) in hand, finding out things like time dilation, length contraction, etc. is a very simple matter. As a preparation for this, just note that the equations are linear, and thus differences of space time coordinates obey the same set of equations :

$$\Delta x' = \frac{\Delta x - v\Delta t}{\sqrt{1-\frac{v^2}{c^2}}} \quad (3.25\text{-a}) \qquad \Delta x = \frac{\Delta x' + v\Delta t'}{\sqrt{1-\frac{v^2}{c^2}}} \quad (3.26\text{-a})$$

$$\Delta y' = \Delta y \quad (3.25\text{-b}) \qquad \Delta y = \Delta y' \quad (3.26\text{-b})$$

$$\Delta z' = \Delta z \quad (3.25\text{-c}) \qquad \Delta z = \Delta z' \quad (3.26\text{-c})$$

$$\Delta t' = \frac{\Delta t - \frac{v}{c^2}\Delta x}{\sqrt{1-\frac{v^2}{c^2}}} \quad (3.25\text{-d}) \qquad \Delta t = \frac{\Delta t' + \frac{v}{c^2}\Delta x'}{\sqrt{1-\frac{v^2}{c^2}}} \quad (3.26\text{-d})$$

Let me quickly illustrate how you would go about discovering all the phenomena that we have covered so far in this chapter from the above equations.

3.5.1.1 Time dilation

Bob's wristwatch shows a time interval of t_0 between two ticks (this must be the proper time - Bob being on the spot at both the ticks). What time interval t will Alice measure? Note that in this case,

$\Delta x' = 0$ and $\Delta t' = t_0$, for the two ticks of Bob's wristwatch. Using (3.26-d) immediately leads to

$$t = \Delta t = \frac{t_0 + \frac{v}{c^2} \times 0}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{t_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

which is our familiar time dilation formula, equation (3.1)!

Before moving on to the next topic, let me warn you about something that most people get confused about at some point of time or the other. In the derivation above I have used the inverse Lorentz transformation - why not use the direct ones? Of course you could have used the direct equations too, getting

$$t_0 = \Delta t' = \frac{t - \frac{v}{c^2}\Delta x}{\sqrt{1 - \frac{v^2}{c^2}}}$$

The trouble is - here Δx is the distance between the two ticks that Alice sees and that is not zero! Putting in $\Delta x' = 0$ in (3.25-a) will immediately give $\Delta x = vt$ (this is obvious directly too - after all, the watch is moving away from Alice with a speed v), which will immediately lead to the correct answer. It is much easier, of course, to use the inverse transformation for this particular case, though - and it is a cardinal sin to mix the two up!

3.5.1.2 Length contraction

Bob carries with him a rod of length L_0. What is the length L that Alice measures for it. Since the rod is at rest with respect to Bob, he will have $\Delta x' = L_0$ for any two events that measure the coordinates of the two endpoints - irrespective of whether they are simultaneous or not! On the other hand, as far as Alice's measurements are concerned, $\Delta x = L$ only if the two measurements are done simultaneously, $\Delta t = 0$. In this example, it is more convenient to use the direct transformation, equation (3.25-a) which gives

$$L_0 = \frac{L - v \times 0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

leading immediately to (3.9)

$$L = L_0 \sqrt{1 - \frac{v^2}{c^2}}$$

3.5.1.3 Relativity of simultaneity

Two events are simultaneous with respect to Alice and she sees them occur at a gap of $\Delta x = l$. The direct transformation equation (3.25-d) immediately gives

$$\Delta t' = \frac{0 - \frac{v}{c^2} \times l}{\sqrt{1 - \frac{v^2}{c^2}}} = -\beta\gamma \frac{l}{c}$$

as the time interval between the two events as seen by Bob.

Deriving the time interval in this way immediately tells us something that the radar method with its one space dimension, could not have told us. The l that determines how far away in time the two events, simultaneous for Alice, will be for Bob - is not the spatial distance between the two events as seen by Alice, but rather, Δx, the component of the separation vector in the x direction (Of course, I have already shown you how to derive the result using Einstein's train example is subsection 3.3.1) !

You may be slightly puzzled by the fact that the x component appears to play a special role in the relativity of simultaneity - after all, aren't all directions in space supposed to be equivalent to each other? A moment's thought will tell you though that the x direction is special - it is the direction of the relative motion of the two friends! Of course, if Bob had been moving away from Alice in any other direction, what would have occurred in the formula above is the component of the separation vector in *that* direction. It is easy to find out the formula that will work for an arbitrary direction of relative motion - all you have to do is rotate the coordinates around in such a way that the new $X - X'$ axes point along the velocity. In this frame, of course, the time interval will be given by $-\frac{1}{c^2}\gamma\,(v\Delta x)$ - but the combination $v\Delta x$ can be written as $\vec{v}\cdot\Delta\vec{r}$, since in this frame $\vec{v}$ has an X component only. Thus the expression in the rotated frame is $-\frac{1}{c^2}\gamma\,(\vec{v}\cdot\Delta\vec{r})$. Now, since this expression involves scalars only (the γ contains v^2, and that is a scalar too) - the expression will stay the same for any rotated coordinate system, including, in particular, the original one! So, in general, if Alice sees two events separated by $\vec{l}$ to be simultaneous, then to bob, who is moving away from Alice with a velocity $\vec{v}$, the separation in time between the two events is

$$\Delta t_{\text{Bob}} = -\gamma\frac{\vec{v}\cdot\vec{l}}{c^2}$$

which is the general formula that we had found out in subsection 3.3.1.

3.5.1.4 The transformation of velocities

We have derived the formula for the addition of velocities as well as the formula for relative velocity using the radar method in section 3.4. I will now show that these follow immediately from the Lorentz transformation equations. As an added bonus, I will now derive the formulae for velocity addition (and relative velocity) if the observed object is moving in an arbitrary direction - not just in the same direction as the relative velocity between Bob and Alice.

Both Bob and Alice measure the spacetime coordinates of their friend Charlie at two events. To Alice, Charlie's velocity in the interval between the two events is given by

$$u_x = \frac{\Delta x}{\Delta t}, \quad u_y = \frac{\Delta y}{\Delta t}, \quad u_z = \frac{\Delta z}{\Delta t}$$

while, according to Bob, it has the components

$$u'_x = \frac{\Delta x'}{\Delta t'}, \quad u'_y = \frac{\Delta y'}{\Delta t'}, \quad u'_z = \frac{\Delta z'}{\Delta t'}$$

From this it is easy to see that

$$\begin{aligned} u'_x &= \frac{\Delta x'}{\Delta t} \times \left(\frac{\Delta t'}{\Delta t}\right)^{-1} \\ &= \frac{\left(\frac{\Delta x}{\Delta t} - v\right) / \sqrt{1 - \frac{v^2}{c^2}}}{\left(1 - \frac{v}{c^2}\frac{\Delta x}{\Delta t}\right) / \sqrt{1 - \frac{v^2}{c^2}}} \\ &= \frac{u_x - v}{1 - \frac{u_x v}{c^2}} \end{aligned}$$

which is the same formula that we found out before.

Working from the Lorentz transformations it is easy to find out the way in which the components of the velocities in the other directions change. You should check that repeating the above steps for u_y or u_z gives the equations

$$\begin{aligned} u'_y &= \frac{u_y \sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{u_x v}{c^2}} \\ u'_z &= \frac{u_z \sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{u_x v}{c^2}} \end{aligned}$$

These are of course, the same set of formulae that you have seen before.

As you can see, the velocity transformations are quite a bit more complicated than the Lorentz transformations themselves - in particular, the components in the directions perpendicular to the motion change in a quite involved manner, while the space coordinates do not change at all!

It turns out that the components of the vector $\vec{U} = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}}\vec{u}$ transforms much more neatly than $\vec{u}$ itself. Here the u in the square root is the magnitude of the vector $\vec{u}$ itself and *not* the relative speed of the two observers. What makes this possible is the rather cute identity

$$\frac{\sqrt{1-\frac{v^2}{c^2}} \times \sqrt{1-\frac{u^2}{c^2}}}{\sqrt{1-\frac{u'^2}{c^2}}} = 1 - \frac{u_x v}{c^2} \tag{3.27}$$

which you should be able to prove for yourself after a bit off (slightly) messy algebra. Check for yourself that this means, for example that

$$\begin{aligned} U'_y &= U_y \\ U'_z &= U_z \end{aligned}$$

Find out the way in which U_x transforms.

One thing that (3.27) shows quite clearly is that if both u and v are less than c, u' must be less than c also! Can you figure this out?

There is, actually, a deeper reason behind why the transformation of $\vec{U}$ is somewhat simpler than that of the more familiar $\vec{u}$. I will tell you about it in a later section where we will study the generalization of our old friends the vectors so that they fit comfortably in the space-time fabric of relativity.

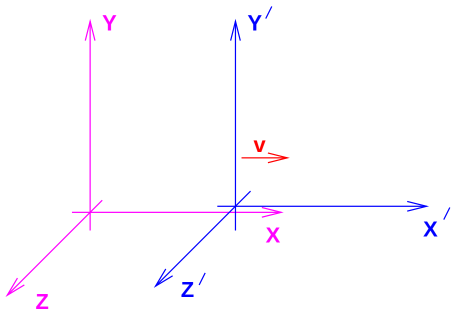
3.5.2 The general Lorentz transformation

The Lorentz transformations that we have been dealing with so far are rather special - they are ones is which the relative velocity of Bob and Alice is along the common X axis, and their coordinate axes are parallel to each other. This is why this is called a **special Lorentz transformation**. However, a Lorentz transformation, in general, is the relationship between the spacetime coordinates as measured by any two inertial observers. And as I have already observed, such observers may move with respect to each other in any direction whatsoever, provided the velocity stays a constant in time. What's more, it is by no means necessary that the two observers use coordinate axes that are parallel to each other. They may just as readily use coordinate axes that are rotated with respect to each other (note - *rotated* and not *rotating*. In the latter case, the observer would no longer be inertial.). Figure 3.5 shows the possibilities.

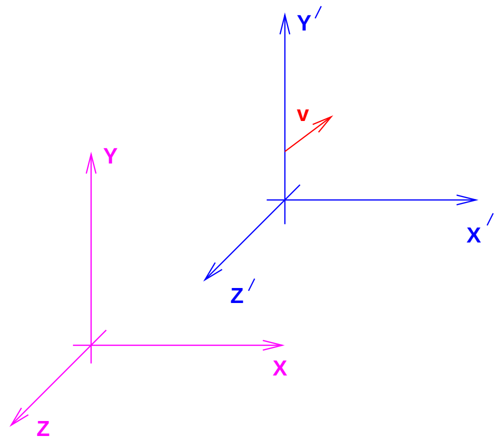
The kind of Lorentz transformations where the two observers use coordinate systems parallel to each other are given a special name - they are called Lorentz **boosts**. As you can easily see, the special Lorentz transformation is actually a special kind of boost - a boost along the X axis.

A more general Lorentz transformation can be easily understood as a boost followed by a rotation. In figure 3.5c, the dashed blue lines show the result of a boost corresponding to the velocity $\vec{v}$. The actual coordinate system for the observer S' is obtained by rotating this dashed coordinate system.

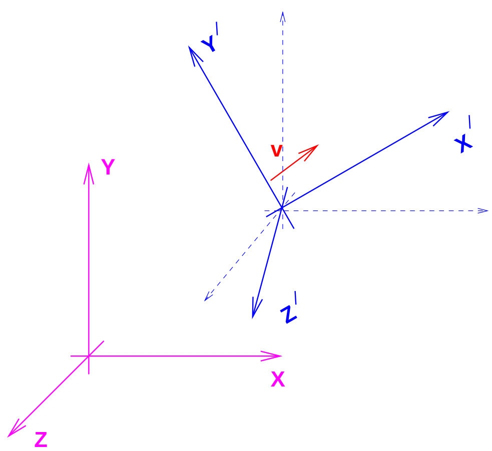
How can we find out how coordinates change under a general boost? You can easily figure this out from what we know of the



(a) Special Lorentz transformation



(b) General Lorentz Boost



(c) General Lorentz transformation

Figure 3.5: Various kinds of Lorentz transformations (a) Special Lorentz transformation - note that the axes of both observers stay parallel to each other, and the relative velocity is along the common $X - X'$ axes. (b) The general Lorentz Boost - the axes are still parallel to each other, but the relative velocity is along some arbitrary direction. Note that the special Lorentz transformation is nothing but a boost along the X axis. (c) The general Lorentz transformation - not only is the relative velocity in an arbitrary direction, but the axes are also rotated with respect to each other.

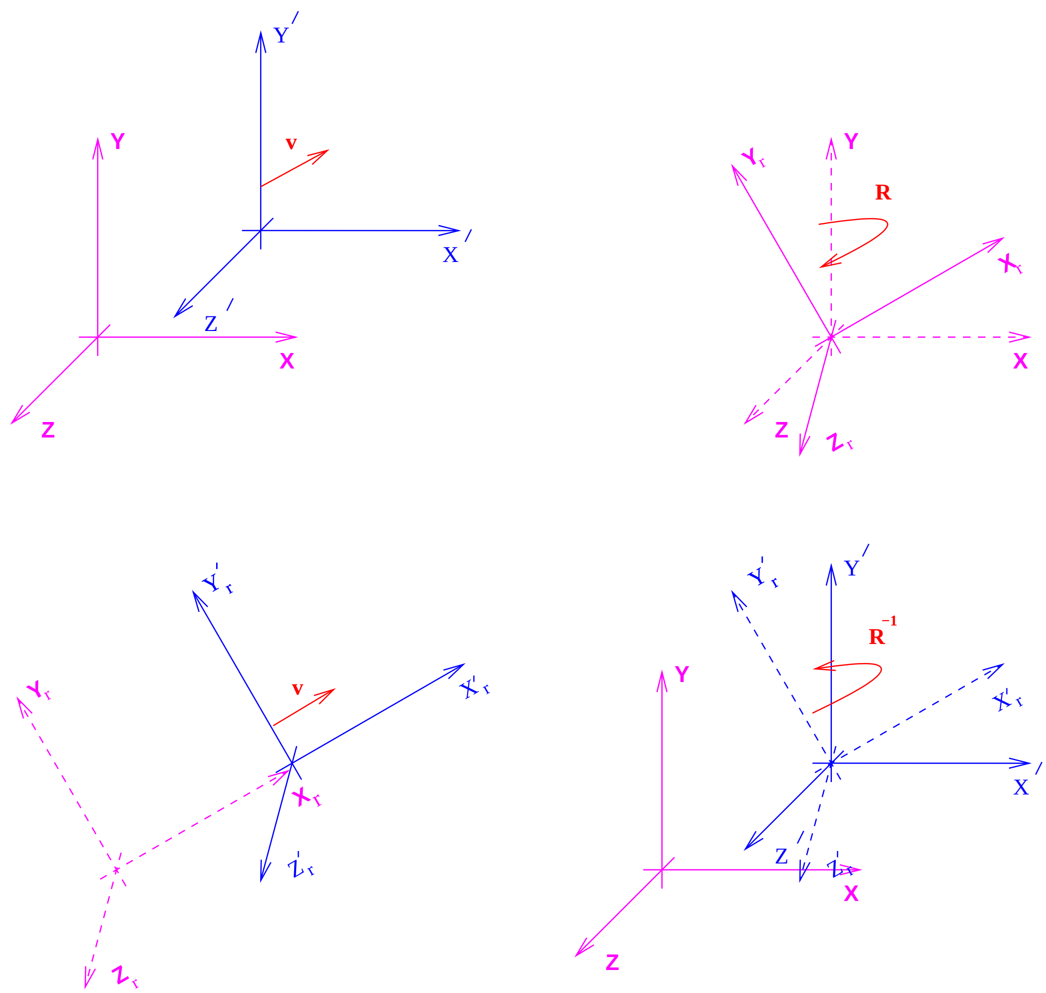


Figure 3.6: A general boost from the special boost.

special case, provided we also know how to work out the way coordinates change under a general rotation. What's needed is shown in figure 3.6. The steps are - first, rotate Alice's coordinate system so that her new X axis (the one denoted X_r in the figure) aligns with the direction of Bob's velocity. Now, use the special Lorentz transformation to boost up to Bob's speed - this gives us the coordinate frame denoted by X'_r, Y'_r and Z'_r. This, though, is not Bob's frame of reference - the sped is right - the orientation is not. to correct for this, we need the final step - carry out the reverse of the rotation in the first step. this leaves us with a coordinate system that is parallel to the original one and is moving with the proper velocity with respect to it - Bob's frame!

All that is fine - and abstract! to understand this better let me show you a concrete example. Instead of trying to figure out the boost in a very general direction (it will be very difficult to check whether the result is correct anyway!) - I will show you how to figure out the equations for Boost along the Z axis.

First, we need a rotation that will take Alice's X axis into her Z axis. There are of course many rotations that can do the job for us - I will use the most obvious - a $90°$ rotation about the Y axis. This gives

$$\begin{aligned} x &\rightarrow x_r = z \\ y &\rightarrow y_r = y \\ z &\rightarrow z_r = -x \\ t &\rightarrow t_r = t \end{aligned}$$

where I have included the (trivial) last line to make the notation consistent. Now, boost to the speed v to get

$$\begin{array}{rcl} x_r & \rightarrow & x_r' = \gamma\left(x_r - vt_r\right) = \gamma\left(z - vt\right) \\ y_r & \rightarrow & y_r' = y_r = y \\ z_r & \rightarrow & z_r' = z_r = -x \\ t_r & \rightarrow & t_r' = \gamma\left(t_r - \frac{v}{c^2}x_r\right) = \gamma\left(t - \frac{v}{c^2}z\right) \end{array}$$

The final step is, of course, carrying out the opposite rotation to the one in the first step.

$$\begin{array}{rcl} x_r' & \rightarrow & x' = -z_r' = x \\ y_r' & \rightarrow & y' = y_r' = y \\ z_r' & \rightarrow & z' = x_r' = \gamma\left(z - vt\right) \\ t_r' & \rightarrow & t' = t_r' = \gamma\left(t - \frac{v}{c^2}z\right) \end{array}$$

So the upshot of it all is

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

If you are comfortable with matrices you should be able to see that the calculation above could be carried out as a matrix product

$$\begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \gamma & 0 & 0 & -v\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{v}{c^2}\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

where the first matrix to be applied is the *rightmost*[4] 4×4 matrix that rotates Alice's X axis to her Z axis. The next to have a go is the matrix in the middle, which is just the one that carries out our special boost. Last to get into the act is the matrix in the front - which is just the inverse of the first rotation. You should carry out the matrix multiplication and check that we do get our final result correct! The beauty of using matrices of course is that the entire calculation of the last page or so can be condensed into a few lines. In the matrix notation, the instructions for finding out the general boost above can be condensed into

$$\mathrm{x}' = R^{-1} L_x R \mathrm{x} \tag{3.28}$$

in an obvious notation.

The above description of how to figure out the transformation equations for a general boost are perfectly correct - but there is, however, a more direct way of figuring out the result for a general boost. Just note that in the special Lorentz transformation equations, all that is special about the x coordinates that it is the component of the position vector of an event that happens to be in the same direction as the relative velocity of Bob and Alice. the fact that y and z do not change under this transformation can be restates as - components of position vector transverse to the relative velocity do not change. So, it pays off to decompose the position vector of the event into two pieces, one parallel to $\vec{v}$ and one perpendicular to it

$$\vec{r} = \vec{r}_{\parallel} + \vec{r}_{\perp} \tag{3.29}$$

[4]Remember, matrices act to their right and thus in a product AB, it is Bthat gets the first shot, not A!

where elementary vector algebra tells you that

$$\vec{r}_{\parallel} = \frac{\vec{r}\cdot\vec{v}}{v^2}\,\vec{v} \tag{3.30}$$
$$\vec{r}_{\perp} = \vec{r}-\vec{r}_{\parallel} \tag{3.31}$$

In terms of this division, the boost becomes

$$\begin{aligned}
\vec{r}'_{\parallel} &= \gamma\left(\vec{r}_{\parallel}-\vec{v}t\right) = \gamma\left(\frac{\vec{v}\cdot\vec{r}}{v^2}-t\right)\vec{v} \\
\vec{r}'_{\perp} &= \vec{r}_{\perp} = \vec{r}-\frac{\vec{v}\cdot\vec{r}}{v^2}\,\vec{v} \\
t' &= \gamma\left(t-\frac{1}{c^2}v\left|\vec{r}_{\parallel}\right|\right) = \gamma\left(t-\frac{\vec{v}\cdot\vec{r}}{c^2}\right)
\end{aligned}$$

Putting this together, we find the final condensed(?!) form

$$\vec{r}' = \vec{r}+\left([\gamma-1]\,\frac{\vec{v}\cdot\vec{r}}{v^2}-\gamma t\right)\vec{v} \tag{3.32-a}$$
$$t' = \gamma\left(t-\frac{\vec{v}\cdot\vec{r}}{c^2}\right) \tag{3.32-b}$$

Check that in the two special cases that we have worked out so far ($\vec{v}$ along the X and the Z axes, respectively) these equations do reduce to the ones we have found out.

This should immediately tell you how to figure out the equations for a general Lorentz transformation. The job is almost done - all that is left is one final rotation - one that will align the coordinate system parallel to Alice's to Bob's actual coordinate system. If I write the action of this rotation as $\mathcal{R}$, I can write down the equations for the general Lorentz transformation as

$$\vec{r}' = \mathcal{R}\left(\vec{r}+\left([\gamma-1]\,\frac{\vec{v}\cdot\vec{r}}{v^2}-\gamma t\right)\vec{v}\right) \tag{3.33-a}$$

$$t' = \gamma\left(t - \frac{\vec{v}\cdot\vec{r}}{c^2}\right) \tag{3.33-b}$$

- as is obvious, the final rotation only changes the position vector and does not touch the time.

Is this, then, the most general way in which spacetime coordinates are related for two inertial observers? The answer, which may come as a bit of surprise to you, is - no! The trouble is, the above equations are homogeneous - $\vec{r} = 0$, $t = 0$ changes to $\vec{r'} = 0$, $t' = 0$. Of course, in our derivation we had assumed that at one instant of time, Alice and Bob were side by side, and they had both set their respective watches to zero at that instant. In general, however, two observers may very well have origins of both time and space translated by constant amounts. Taking this into account gives us the most general transformations that can relate the spacetime coordinate measurements of two inertial observers as

$$\vec{r'} = \mathcal{R}\left(\vec{r} + \left([\gamma - 1]\,\frac{\vec{v}\cdot\vec{r}}{v^2} - \gamma t\right)\vec{v}\right) + \vec{r}_0 \tag{3.34-a}$$

$$t' = \gamma\left(t - \frac{\vec{v}\cdot\vec{r}}{c^2}\right) + t_0 \tag{3.34-b}$$

where $\vec{r}_0$ and t_0, are obviously constants that represent the space and time translations. These final transformations are called the Poincaré transformations in honour of the great French polymath Henri Poincaré. There are also known, for obvious reasons as inhomogeneous Lorentz transformations. In summary, then, the Poincaré transformations, which express the most general connection between two inertial observers, comprise of boosts, rotations and spacetime translations.

3.5.3 Why use the radar method?

The foregoing may have prompted to ask you the following question - "If everything can be derived from the Lorentz transformations, why bother to go through the lengthier route of the radar calculus at all?". In my defense I will put forward the following points

- The direct derivation of the Lorentz transformations is not as simple as the radar method to understand (although as we will see, it is not very difficult, either!).
- More importantly, learning relativity from the LT equations only leaves the impression that the whole thing is essentially a mathematical sleight of hand. The radar method has, in my honest opinion, a much more physical feel to it.
- Most importantly, the radar method emphasises the central message of STR - pay careful attention to how we measure space and time!

3.6 The invariant interval

A side result that I had shown you while deriving the Lorentz transformation equations is (3.22), which states that the quantity $t^2 - \frac{x^2}{c^2}$ is the same for both Alice and Bob. Since the measurements the two make of lengths perpendicular to their relative motion are bound to be the same, we can generalize (3.22) to 3 space dimensions directly to get

$$t^2 - \frac{x^2 + y^2 + z^2}{c^2} = t'^2 - \frac{x'^2 + y'^2 + z'^2}{c^2}$$

a result that is often written as

$$c^2t^2 - r^2 = c^2t'^2 - r'^2$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is the distance between the point where the event occurred and the origin. Since the differences of space-time coordinates of two events transform in the same way as the coordinate differences themselves, it is easy to see that the **interval** between two events, defined by[5]

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \qquad (3.35)$$

is the same for Bob as for Alice,

$$\Delta s^2 = \Delta s'^2 \qquad (3.36)$$

You may have a slightly uncomfortable feeling that in our derivation we have considered a rather special case, where Bob moves away from Alice along the common X axis, while they both make use of coordinate axis that are parallel to each other's. A moment's thought will tell you that you can always rotate each friend's coordinate system so that their common X axis gets aligned along the direction of their relative velocity. If we use these rotated coordinate systems, the proof of (3.36) goes through. During the rotations, the time interval between two events stay the same, and while the individual components $\Delta x, \Delta y$ and Δz change, the sum of their squares do not! This means that no matter what coordinate systems the two friends use and which direction their relative velocity points in, (3.36) remains valid for them!

[5]Here, I am abusing notation slightly. Δt^2 stands for $(t_2 - t_1)^2$ and *not* $t_2^2 - t_1^2$, and so on! Of course, I could have written the more precise $(\Delta t)^2$ - but it seems that the usage here is considered to be standard by most authors.

Given the importance that the interval has in relativity, you would be surprised if there was no other way of deriving this result! So let me give you a more direct proof of the invariance of the interval. For this, just consider two special events - a flash of light being sent out from some point in space, and reaching some other point a while later. For Alice, the time gap between these two events must be related to the distance between them by $c\Delta t = \Delta r$ - which is just the statement that light moves with a speed c as seen by Alice. So, for Alice we have $\Delta s^2 = 0$ for these two events. Since Bob will also see the flash travel with the same speed, a very similar argument will tell you that $\Delta s'^2 = 0$, too. Again, if the interval between any two events is zero as seen by Alice, a light signal can start from the earlier of the two events and get to the position of the latter event just when it occurs. Thus the interval must have the property that whenever Δs is zero, $\Delta s'$ must be zero too. Now, the homogeneity of space tells us that the relation between the space time coordinates as seen by the two friends must be linear (If this seems too cryptic to you, don't worry - I will give you a much more detailed explanation of this when I give you a direct derivation of the Lorentz transformation equations in subsection 5.1). Also, since if the two events occur at the same time *and* the same place for Alice, Bob must also agree that they are coincident - the relation between space-time coordinate differences must be homogeneous as well. This, couple with the fact that they vanish together, means that the quadratic expression for Δs^2 must be related to $\Delta s'^2$ by a relation given by

$$\Delta s'^2 = \lambda \Delta s^2$$

where λ is a dimensionless quantity that is independent of the space time coordinates. However, at this stage, it is quite possible that λ depends on the dimensionless ratio $\beta = \frac{v}{c}$, *i.e.* $\lambda = \lambda(\beta)$.

All that remains to be done now, is to convince you that λ must be 1 - irrespective of what the relative speed of the two friends is. For this, just consider two events that Alice sees occurring at the same time, but at two different points located along a line perpendicular to the direction in which Bob is moving (In our special choice of coordinates, this may mean that $c\Delta t = \Delta x = \Delta y = 0$, while $\Delta z \neq 0$). As I have shown you before, these two events will be simultaneous to Bob, too. Couple this with the fact that lengths perpendicular to the direction of motion stay unchanged between observers, and it follows that for these two events, at least, we must have

$$\Delta s^2 = \Delta s'^2 \neq 0$$

which immediately tells us that the factor λ must be one - leading directly to the invariance of the interval.

You may object to the argument I have given you just now on the grounds that it uses too many physical results. For the doubters, I will give another argument, this time lifted straight out of Landau and Lifschitz's classic *The classical theory of fields*[6]. Consider a third friend Charlie, moving with a velocity $\vec{u}$ as seen by Bob. Let's denote his speed, as seen by Alice to be $\vec{w}$. In this case, Charlie's value for the interval between the two events will be

$$\Delta s''^2 = \lambda\left(\frac{u}{c}\right)\Delta s'^2 = \lambda\left(\frac{u}{c}\right)\lambda\left(\frac{v}{c}\right)\Delta s^2$$

which must match $\lambda\left(\frac{w}{c}\right)\Delta s^2$. This means that the function $\lambda\left(\beta\right)$

[6]In case you are wondering what the invariance of the interval is doing in a text with a name that contains "classical theory" - here the word classical means anything that does not involve quantum mechanics - and this includes STR as well as GTR!

must have the property that

$$\lambda\left(\frac{u}{c}\right)\lambda\left(\frac{v}{c}\right)=\lambda\left(\frac{w}{c}\right)$$

Before you start out on a search for such a function, let me point out that in general the relative speed of Charlie with respect to Alice must depend on the angle between $\vec{u}$ and $\vec{v}$. So the right hand side of the equation above depends on this angle, while the left hand side does not depend on it at all! The only way in which this can be true is if λ is a constant - and this means that $\lambda^2=\lambda$, leading to $\lambda=1$.

Let me stress once again that this proof is very general - all that is involved is moving from one inertial observer to another. Perhaps this is flogging a dead horse, but let me show you one more proof of the invariance of the interval. In subsection 3.5.2 I showed you that the most general such transformation is one that involves a boost in a general direction followed by a rotation. Indeed, you can figure out the transformations from the special Lorentz transformations, preceded and followed by rotations. I have already shown above that the interval stays the same under a special Lorentz transformation. Rotation does not change either time or length, leaving the interval invariant. Thus, it is easy to see that the interval does not change for a general Lorentz transformation.

The invariance of the interval goes through in the case of Poincaré transformations, too. Note that since the constant spacetime translations cancels out, both $\Delta\vec{r}$ and Δt for the Poincaré transformations is the same as that for the Lorentz transformations - and that is all you need for the interval to be invariant.

A word on the notation. I have called $c^2\Delta t^2-\Delta r^2$ the interval. From the experience that you have in ordinary geometry, where the *square root* of $\Delta x^2+\Delta y^2+\Delta z^2$ is the spatial length, you may think that is better to call the square root of this quantity the interval. The trouble is, of course, that the invariant combination above need not be positive - it is negative for two events that occur far enough apart in space so that Δr exceeds $c\Delta t$. I will play it safe (most people do) and keep on calling the expression in (3.35) the interval.

Another point about the notation - instead of using $c^2\Delta t^2 - \Delta r^2$ the interval, I could just as well have reserved this name for the quantity $\Delta r^2 - c^2\Delta t^2$. Which one to use is primarily a matter of preference and both conventions are in use. In these lectures I will use the former exclusively - but do check the convention that is being used when you read any other book on STR.

3.6.1 Splitting up spacetime

The fact that the interval between two events is the same for all inertial observers lead us to the notion that pairs of events can be classified in a very significant way using the interval. Two events that have a positive interval between them are said to be separated by a **timelike** interval, while if $\Delta s^2$is negative, the interval is called **spacelike**. The nomenclature is obvious from the way the interval is defined - for timelike intervals, the separation in time is larger than that in space (if both time and space are measured in the same units - leading to $c = 1$), while for spacelike intervals the spatial part dominates. Finally events for which $\Delta s = 0$ are said to be separated by a **light-like** or **null** interval, for obvious reasons. note that the fact that any two inertial observers measure the same interval leads to everybody agreeing on whether the separation between two events is timelike, spacelike or null - even though they disagree about the actual values of the space-time coordinates of the two events! Very often, I will follow the general practice of saying “two events are timelike” instead of the more long-winded “two events are separated by a timelike interval”.

You may be wondering why I am harping so much on the sign of the interval - and dividing up event pairs into just three categories depending on whether it is positive, negative or zero - when not only just the sign but the entire value of the interval is the same for all observers. Why not divide spacetime into an infinite number of classes - each class consisting of event pairs separated by the same value of the interval? All inertial observers will certainly agree on whether a particular pair of events belong to a particular class. In the following paragraphs I will give you a few reasons why the division of spacetime events into light-like, spacelike and timelike pairs is physically significant. However, the real physical significance of this division will emerge only when we talk about causality in section 3.9.

It directly follows from the definition of the interval that two events that occur at the same place (but at different times) to Alice, have a positive interval between them. Since Bob must measure the same positive interval between them, even though he will see them to occur at different places, it follows that such events are timelike for everybody. Again, if Alice sees two events occur simultaneously but at different places, then the two events are spacelike for everybody. Turning this over in its head, it is easy to argue that if Alice sees a timelike interval between two events, then Bob will never see them to be simultaneous - or if Alice sees a spacelike interval between two events then Bob will never see them as occurring at the same place.

Again, if Alice sees two events have a timelike interval - it is possible for Bob to move with such a speed that the two events will occur at the same place with respect to him. Note that this means that Alice will see Bob have a speed given by $\frac{\Delta r}{\Delta t}$ - which is less than the speed of light as $c\Delta t > \Delta r$ in this case. On the other hand, if Alice sees a spacelike interval between them, it is possible for Bob to move fast enough so that the two events are simultaneous with respect to him. As you can easily figure out from (**??**), Bob's velocity for this case will have to obey

$$\gamma \frac{\vec{v} \cdot \Delta \vec{r}}{c^2} = \Delta t$$

which means that the component of Bob's velocity parallel to $\Delta \vec{r}$ must be

$$v_{\parallel} = \frac{c}{\gamma} \times \frac{c\Delta t}{\Delta r}$$

Since both γ and the ratio $\frac{\Delta r}{c\Delta t}$ are larger than 1 (the first, always for a nonzero speed, and the second because the interval is spacelike) - this component is less than c. Thus, it is possible for Bob to move with such a speed.

3.6.2 Everything from the interval!

As we have seen a lot of times before - there are many apparently different ways to arrive at the same physical conclusion in relativity. In the following, I will show you how to exploit the invariance of the interval to derive things like time dilation etc.

3.6.2.1 Time dilation

Consider two clocks on Bobs watch. The time interval that he measures between them is the proper time interval $\Delta\tau$ of these two events, while the spatial interval is, of course, 0. What does Alice measure the time interval Δt between the two ticks? The invariance of the interval tells us immediately that

$$c^2\Delta\tau^2 = c^2\Delta t^2 - \Delta r^2$$

where Δr, the spatial separation between the two ticks as seen by Alice is related to Δt by $\Delta r = v\Delta t$. This means

$$\Delta\tau^2 = \Delta t^2\left(1 - \frac{1}{c^2} \times \frac{\Delta r^2}{\Delta t^2}\right) = \Delta t^2\left(1 - \frac{v^2}{c^2}\right)$$

which immediately leads us to the familiar time dilation formula.

What about the other physical results like length contraction, relativity of simultaneity etc.? Well, it is possible to derive these results using the invariance of the interval - but the arguments are more involved than the very simple one that gives us time dilation. Anyway, as we will see in subsection 5.1, you can actually derive the Lorentz transformations themselves from the invariance of the interval - and that, of course, allows you to derive all these results rather straightforwardly.

3.7 Acceleration in relativity

3.7.1 The transformation law for accelerations

We have already seen how space-time coordinates and velocities change when we move over from one inertial frame to another. It is now the turn of accelerations. This was really simple in the good lod days of Galilean transformations - the accelerations measured by all inertial observers were, simply, the same (as long as we stick to frames that are not rotated with respect to each other)! Thing are quiet a bit more complicated under the Lorentz transformation.

Let us now consider a special Lorentz transformation from the frame S to the frame S'. As we have already seen, the volcity transforms as

$$
\begin{aligned}
u_1' &= \frac{u_1 - v}{1 - u_1 v/c^2} \\
u_2' &= \frac{u_2\sqrt{1 - v^2/c^2}}{1 - u_1 v/c^2} \\
u_3' &= \frac{u_3\sqrt{1 - v^2/c^2}}{1 - u_1 v/c^2}
\end{aligned}
$$

Now, the acceleration components measured by S and S' are, simply $a_1 = du_1/dt$ and $a_1' = du_1'/dt'$ and so on, respectively. Thus

$$
\begin{aligned}
a_1' &= \frac{du_1'}{dt'} = \frac{du_1'/dt}{dt'/dt} \\
&= \frac{\frac{d}{dt}\left(\frac{u_1 - v}{1 - u_1 v/c^2}\right)}{\frac{d}{dt}\left(\frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}\right)}
\end{aligned}
$$

Carrying out the differentiation for this as well as the other two components yield the acceleration transformations

$$
a_1' = \frac{\left(1 - v^2/c^2\right)^{3/2}}{\left(1 - u_1 v/c^2\right)^3} a_1 \tag{3.37-a}
$$

$$
a_2' = \frac{\left(1 - v^2/c^2\right)}{\left(1 - u_1 v/c^2\right)^3}\left[a_2 - \frac{v}{c^2}\left(u_1 a_2 - u_2 a_1\right)\right] \tag{3.37-b}
$$

$$
a_3' = \frac{\left(1 - v^2/c^2\right)}{\left(1 - u_1 v/c^2\right)^3}\left[a_3 - \frac{v}{c^2}\left(u_1 a_3 - u_3 a_1\right)\right] \tag{3.37-c}
$$

for the special Lorentz boost.

I admit that these transformations look rather intimidating - even for the case of rectilinear motion, where only the first equation is involved. There is one case, however, where the 1-D version

$$a' = \frac{\left(1 - v^2/c^2\right)^{3/2}}{\left(1 - uv/c^2\right)^3} a_1 \tag{3.38}$$

simplifies a lot. This is the case where the frame S happens to be the *Instantaneously Comoving Frame* (ICF) of the particle in question. This is the frame in which the particle is at rest at a given instant of time - so that $u = 0$. Remember, an ICF is *not* a frame fixed to the accelerating particle - that, of course, would be a non-inertial frame! Thus, though the velocity components vanish in this frame, they do so only for a particular instant! Thus, the acceleration components are, in general, nonzero in the ICF. Of course, if S is an ICF, then the velocity v of S' with respect to S is nothing other than the speed $-u'$, where u' is the velocity of our particle in the S' frame. Thus, in this special case we have

$$\left(1 - \frac{u'^2}{c^2}\right)^{-3/2} a' \quad = \quad a_0 \tag{3.39}$$

where a_0 is the acceleration of the particle in the ICF. This, of course, is a frame independent quantity by definition, and is called the *proper acceleration* of the particle. Note that we had no necessity of such a concept in classical physics - acceleration being the same in all inertal frames. Since the frame S' is arbitrary we see that for rectilinear motion, we must have

$$\gamma^3 \left(u'\right) a' = \gamma^3 \left(u''\right) a'' = \dots$$

3.7.2 Uniformly accelerated motion in STR

One of the first problems that we study in classical mechanics is that of uniformly accelerated motion. Every high school kid knows that

$$a \equiv \frac{d^2x}{dt^2} = \frac{du}{dt} = \alpha, \qquad (\alpha = \text{constant})$$

has the solution

$$\begin{aligned} u &= u_0 + \alpha t \\ x &= x_0 + u_0 t + \frac{1}{2}\alpha t^2 \end{aligned}$$

where u_0 and x_0 are the values of u and x, respectively, at $t = 0$.

Of course, if $a = \alpha =$ contsant, we will always get the above solution - be it classical physics or relativity! However, the above notion of uniform acceleration does not make much sense in relativity. After all, even if the acceleration were a constant in one inertial frame, the transformation law (3.38) tells us that that in another frame will be time dependent (note that the acceleration transformation involves, in addition to the acceleration, also the velocity of the particle in the frame S - which is where the time dependence comes from)! Thus, our old notion of uniform acceleration will be valid, if at all, in one inertial frame only - and as such, cannot have any deep physical significance.

On the other hand, we have met a notion of the acceleration which is frame independent in the last section - the proper acceleration! If the proper acceleration of a body is a constant in one frame, then it is the same constant in all frames (after all, it is the acceleration of the body as observed from one particular frame - its ICF). So, the notion of uniform acceleration that makes sense in relativity is that of *uniform proper acceleration.* In what follows we will investigate this kind of uniformly accelerated motion.

It might appear that the simplest way to solve the equation $a_0 = \alpha$ is to solve it in the ICF - where the soltion will be just the same as that in high school physics. This is wrong, however! The point is - the ICF of an accelerated body is not a single frame - you have a new ICF every instant of time! To solve this problem, you have to transform the uniform proper acceleration to a fixed inertial frame (which we will call the frame S in this case). According to (3.39) this gives the equation

$$\left(1-\frac{u^2}{c^2}\right)^{-3/2}\frac{du}{dt}=\alpha \tag{3.40}$$

which can be integrated to yield

$$\frac{u/c}{\left(1-u^2/c^2\right)^{1/2}}=\frac{\alpha t}{c}$$

where we have assumed that the frame S is the frame in which the particle is at rest at $t = 0$. This can be simply rearranged to give

$$u=\frac{\alpha t}{\sqrt{1+\left(\alpha t/c\right)^2}} \tag{3.41}$$

Let us compare this solution for the speed of the particle with the classical result $u = \alpha t$. As you can see, our solution is very close to αt when this speed is much smaller than that of light. This is, of course, only to be expected! However, the deviations show up as the speed gets closer to that of light. What is gratifying is the fact that though the classical result can increase without bound, the expression in (3.41) can never exceed the speed of light!

A simple integration gives the displacement x as

$$x = \frac{c^2}{\alpha}\sqrt{1 + \left(\frac{\alpha t}{c}\right)^2} \tag{3.42}$$

where we have chosen $x = c^2/\alpha$ at $t = 0$ to simplify the expression slightly. For small times this does give us the "parabolic" behavior, $x \approx c^2/\alpha + \alpha t^2/2 + \ldots$, as expected. However, it is easy to see that the world line of the particle, as seen from the S frame is described by the hyperbola

$$x^2 - c^2 t^2 = \frac{c^4}{\alpha^2} \tag{3.43}$$

This is why uniformly accelerated motion in relativity is also often called *hyperbolic motion*.

As can be easily seen, for large times t, the hyperbola approaches $x \to \pm ct$ asymptotically. This makes eminent sense, since the spped approaches $\pm c$ for $t \to \pm\infty$. However, this also throws up a surprise - a light photon starting out from the origin at $t = 0$ (which is when our uniformly accelerated particle was momentarily at rest at $x = c^2/2$) will never catch up with the particle! This, in spite of travelling faster than our particle at all times! In general, light starting more than a distance c^2/α away from our particle when it was momentarily at rest will never catch up with it.

To put the situation in perspective, let us try some numerical estimates. Imagine a rocket on an interstellar journey. To maintain optimum comfort, the proper acceleration of the rocket (which is what an astronaut riding in it will feel) is kept at a comfortable 1 g ≈ 10 ms^{-2}. How long does it take for our rocket to reach a speed of $0.999\,c$? It is

$$\frac{c}{\alpha} \times \frac{0.999}{\sqrt{1 - 0.999^2}} \approx 6.7 \times 10^8 \text{ s} \approx 21 \text{ years!}$$

How long will it take for our astronaut to reach the center of our galaxy - roughly 2×10^{20} m away? Since the time taken here is very large, we can safely approximate x by ct and get the huge time

$$\frac{2 \times 10^{20}}{3 \times 10^{8}} \text{ s} \approx 6.7 \times 10^{11} \text{ s} \approx 21000 \text{ years!}$$

Surely, such an attempt is completely futile for us, mortal men?!

There is hope for our interstellar traveller, though! His journey does take 21000 years for us stay at homes - but remember that for most of that time he has travelled at a very large speed with respect to us - very nearly the speed of light! The immense time dilation caused by this makes him age a lot less!

Let's work out the proper time elapsed when our clock reads a time t. Since we have already found out u as a function of time, it is easy to integrate $d\tau = dt\sqrt{1 - u^2/c^2}$ to find out the proper time elapsed in terms of t:

$$\begin{aligned} \tau &= \int_0^t dt' \sqrt{1 + \left(\frac{\alpha t'}{c}\right)^2} \\ &= \frac{c}{\alpha} \ln\left(\frac{\alpha t}{c} + \sqrt{1 + \left(\frac{\alpha t}{c}\right)^2}\right) \end{aligned} \tag{3.44}$$

This tells us that the 21000 years time that we observe the journey to the center of the galaxy to take corresponds to a proper time of only about 10 years! So our interstellar traveller can make the journey and return in about two decades of his time - only to find all remains of civilization having vanished in the more than 42000 years that has elapsed on earth during his journey!

It may be instructive to write down the expressions for t and x in terms of the proper time τ. You should be able to show easily that

$$t = \frac{c}{\alpha}\sinh\left(\frac{\alpha\tau}{c}\right) \quad (3.45\text{-a})$$

$$x = \frac{c^2}{\alpha}\cosh\left(\frac{\alpha\tau}{c}\right) \quad (3.45\text{-b})$$

3.8 Charting out spacetime

Let's take a second look at the device that we have been using all the time - the spacetime diagram. So far, we have plotted the spacetime diagrams with a bias. They are all diagrams referring to Alice as the observer. Given the democracy among all inertial observers, it is natural to enquire about the shape this diagram takes for Bob.

The first things that Alice would plot in a spacetime diagram are, obviously, the X and T axes. The T axis corresponds to $x = 0$ - it is, as I have already stressed earlier, nothing but Alice's own worldline. The X axis, on the other hand, cannot be the worldline of any moving particle. It is, of course, the line $t = 0$ - the locus of all events simultaneous with Alice's wristwatch showing a reading of 0!

Any kid who has used a sheet of graph paper knows that apart from the two axes themselves, a great deal of utility lies in the coordinate grid. In Alice's spacetime diagram , this would mean two sets of straight lines. First - a set of equi-spaced straight lines parallel to the T axis. Physically, these are the worldlines of a set of particles that are at rest with respect to Alice, spaced out at a regular interval from her. The second set are the equi-spaced lines parallel to the X axis. In our physics based language, each one of these is the locus of a set of simultaneous events - those simultaneous with each successive tick of Alice's wristwatch.

As you can see - worldlines and loci of simultaneous events play a very basic role even in the simplest of spacetime diagrams! Now that we know that Bob does not agree with Alice about which events are simultaneous, it stands to reason that for Bob, the X axis and the coordinate lines parallel to it will be made up of a different set of events[7].

To see what Bob's coordinate grid will look like, then, we must first of all find out the locus of all events that are simultaneous, according to him, to a particular tick on his wristwatch. I will now show you how to do this using school level geometry. For simplicity, I will specialize to the case of events simultaneous with Bob's wristwatch showing 0. Figure 3.7 shows Bob sending out a light signal from event P[8], which bounces at event R and returns to him at Q. If P and Q are equidistant from the origin O, then our fundamental rule, equation (2.2-a), says that the event R is simultaneous with O. Let's join the line OR. also let's extend the line QR to meet Alice's axes at S and T, respectively.

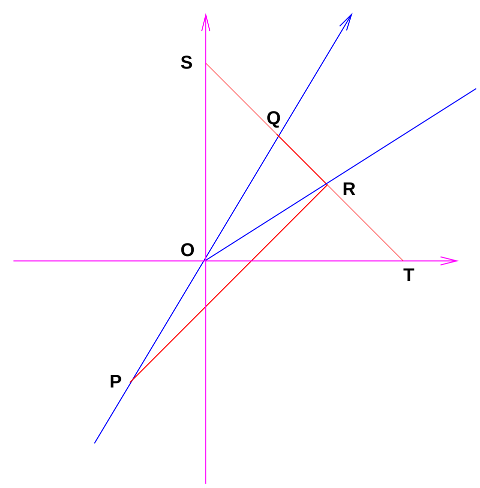


Figure 3.7: Finding Bob's X axis

Now, since light rays make angles of $45°$ with the axes, the $\angle PRQ$ is a right angle. A simple theorem that you studied back in your school days should tell you that this means that the point R lies on the circumference of a circle that has PQ as a diameter, and hence O as its center. Thus ΔOPR is isosceles - leading to $\angle OQR = \angle ORQ$. Since the exterior angle of a triangle is equal to the sum of the two

[7]We already know that Bob's T axis is different from Alice's - it is, after all, his worldline!

[8]To be more precise, I should say - the event that is represented in the spacetme diagram by the point P rather than the event P.

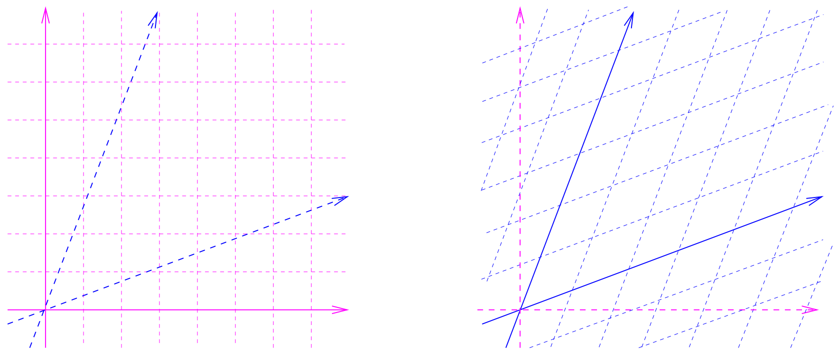


Figure 3.8: Spacetime coordinates according to Alice and Bob. Alice's coordinate lines are in magenta, Bob's in blue.

other angles, it is easy to see that

$$\begin{aligned} \angle TOR &= \angle ORQ - \angle OTR \\ &= \angle OQR - \angle OSQ \\ &= \angle SOQ \end{aligned}$$

Note that this relation must hold for all points R that depicts events that are simultaneous with O as seen by Bob. This means that Bob's X axis is the straight line OR that makes the same angle with Alice's X axis that his time axis makes with Alice's! I leave it as an exercise for you to show that all Bob's lines of simultaneous events turn out to be straight lines parallel to this one. Thus, one half of Bob's coordinate grid consists of equi-spaced straight lines, each inclined to Alice's X axis at the same angle as his worldline is inclined to her time axis.

As for the other half of Bob's coordinate grids - they can obviously be thought of as the worldlines of particles at rest with respect to Bob, placed at equal intervals with respect to him. So, this grid consists of a set of equi-spaced straight lines each parallel to Bob's worldline. Figure 3.8 shows Bob's space and time axes in relation with Alice's coordinate grid on the left. On the right I have drawn Bob's coordinate grid - where I have also shown Alice's axes for reference.You can graphically see that which two events are simultaneous and which are not is dependent on the observer from the fact that Bob's space axis is tilted with respect to Alice's.

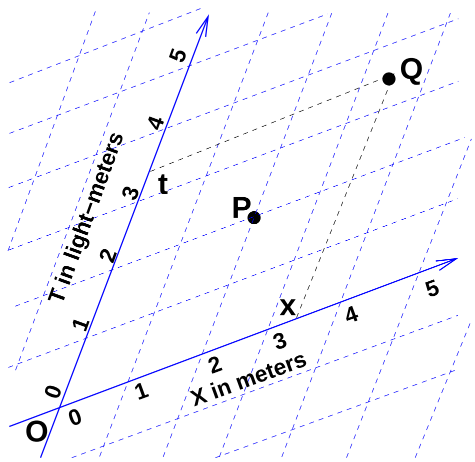


Figure 3.9: Using a non-perpendicular coordinate grid

Figure 3.8 may appear to reveal an asymmetry between the two friends. Alice's space and time axes are nice and perpendicular to each other - the way that we have always learned the axes of a graph should be. Bob's, however, are tilted at a crazy angle to each other! Why this difference?

The point is - what I've drawn for you is really Alice's spacetime diagram - I've just incorporated Bob's spacetime axes in it. You could just as well have drawn Bob's spacetime diagram directly - with axes perfectly perpendicular to each other. In that diagram, as can be guessed, it is Alice's axes that will turn out to be the tilted ones.

You may find the fact that Bob's spacetime axes are not at a right angle slightly discomfiting. How does one use such a set of axes to find the coordinates of any particular event? The coordinate grid, of course, makes it easy - all you have to do is to find which two lines of the coordinate grid intersects at the event - they will immediately tell you what values of x and t the event occurs at. For example, the point P in figure has $x = 2$ m and $t = 2$ l-m[9]. What

[9]Here l-m stands for a light-meter, the time in which light travels a distance

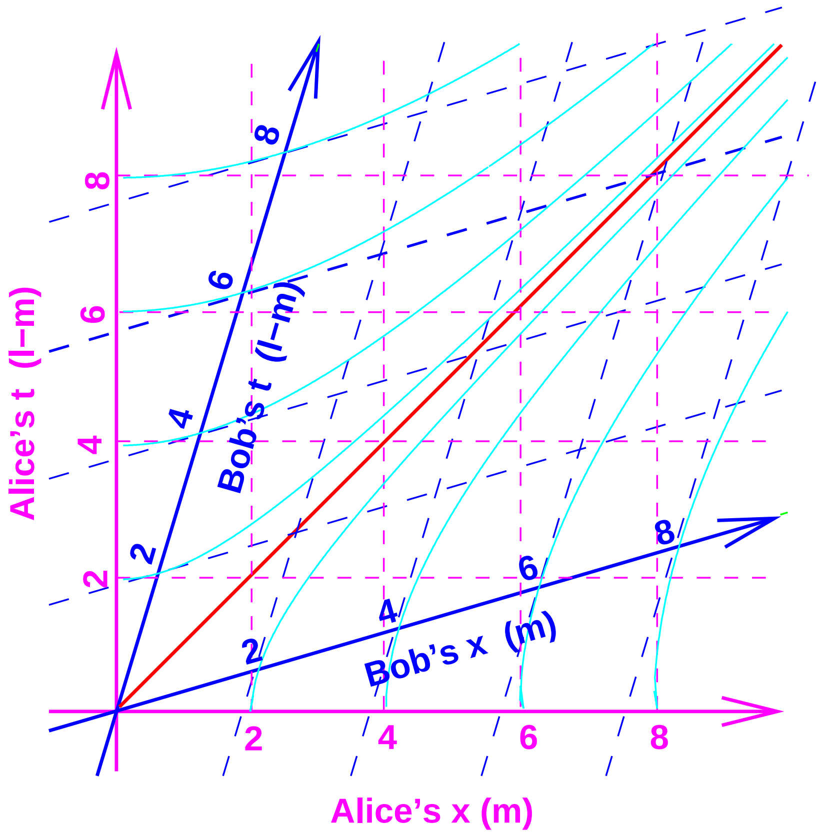


Figure 3.10: Alice and Bob's spacetime coordinates

will we do for events like Q that are not at the corners of the grid? In that case, of course, the grid gives us only an approximate way of locating the event. As you can see, the event Q has x somewhere in between 3-4 m, and the time at which it occurs is also in the range 3-4 l-m. You can certainly go in for a finer grid, which will give you a better approximation. To get a more precise reading, all you have to do is simple - draw lines parallel to the x and t axes from the point Q. Where these cut the t and x axes, respectively, gives the time and position of the event Q.

of 1 m.

Once we have charted out spacetime coordinates for Bob as well as Alice, we are very close to being able to work out how the spacetime coordinates of any given event as measured by the two friends are related to each other. A very simple question needs to be answered before this can be carried out in practice - just how big must the spacing between the coordinate lines of Bob be? The simplest way of seeing the answer is to note that both friends agree about the value of the invariant quantity $x^2 - c^2t^2$. So, a line of events for which $x^2 - c^2t^2 = l^2$ will translate into the line $x'^2 - c^2t'^2 = l^2$for Bob. This line is, of course, a hyperbola. The hyperbola cuts Alice's X axis at $x = l$, and it cuts Bob's X axis at $x' = l$. A set of these hyperbolas, for equally spaced values of l will tell us where to locate both Alice and Bob's constant x lines. Similarly, the hyperbola $x^2 - c^2t^2 = -c^2\tau^2$ cuts the two friends' T axes at $t = \tau$ and $t' = \tau$, respectively. Using these hyperbolas (or rather, the points where they intersect the T axes) we can draw the complete coordinate grid for both friends together. This is what I have drawn in figure 3.10.

As you can see from the figure, the scales used in the two set of axes are quiet different. Figuring out the ratio between the two scales is a rather simple job of elementary coordinate geometry. Bob's T axis, is, of course, his worldline - which, in terms of Alice's coordinates, is the line $x = \beta(ct)$. It is easy to show that his X axis, which is the reflection of this line in the $x = ct$, is given by $x = \frac{1}{\beta}(ct)$. The hyperbola $x^2 - c^2t^2 = l^2$ will cut Bob's X axis at the points $x = \pm\frac{l}{\sqrt{1-\beta^2}}$ (you should have been able to deduce this from what you know about length contraction) and $ct = \pm\frac{\beta}{\sqrt{1-\beta^2}}l$. In Alice's space-time diagram, the distance between these points and the origin is, of course, $\sqrt{\frac{1+\beta^2}{1-\beta^2}}l$. However, this is precisely the length that Bob will call l. This tells us that Bob's unit of length is longer than that used by Alice by a factor of $\sqrt{\frac{1+\beta^2}{1-\beta^2}}$. You should be able to

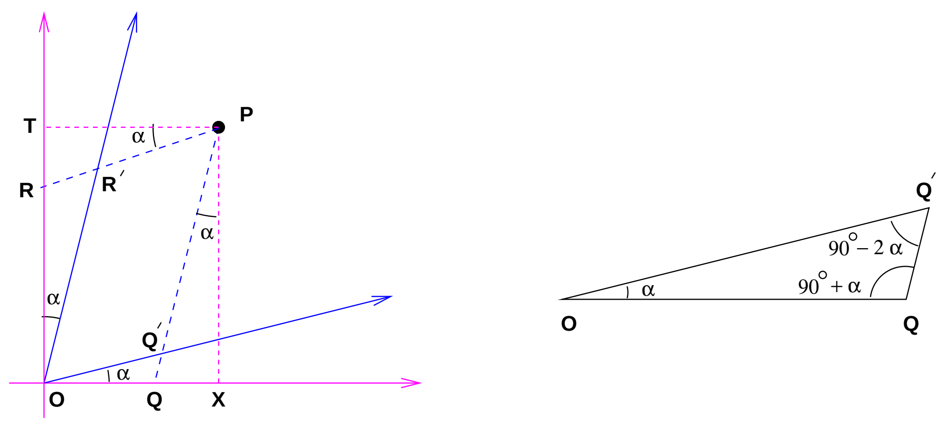


Figure 3.11: Deriving the Lorentz transformations from the spacetime diagram

show quiet easily that the scale factor is the same between the two friends' time axes. Anyway, this should have been obvious from the fact that the line showing the propagation of light is equally inclined to Bob's axes, just as it was to Alice's axes.

Armed with a complete description of the spacetime diagram, we are now in a position to derive the Lorentz transformations themselves from it! The relevant diagram is figure 3.11. To find the coordinates of the point P, all Alice has to do is drop perpendiculars PX and PT to her X and T axes, respectively. Of course, $x = \overline{OQ}$ and $ct = \overline{OT}$. As for Bob, he has to complete the parallelogram $OQ'PR'$, where OQ' and OR' are along his X and T axes. Then OQ' and OR' will represent his measurements of x' and t' - once account is taken of the difference in scale between the two friends. I have extended the lines PQ' and PR' to cut Alice's axes at Q and R, respectively.

Now, from elementary geometry it follows that

$$\angle QPX = \angle TOR' = \angle TRP = \angle Q'OX = \alpha = \tan^{-1}\frac{v}{c} = \tan^{-1}\beta$$

This tells us that $\overline{QX} = \overline{PX}\ \tan\alpha = \beta ct = vt$ and $\overline{TR} = \overline{PT}\ \tan\alpha = \beta x$. Thus, $\overline{OQ} = x - vt$ and $\overline{OR} = ct - \beta x$. What I need to figure out Bob's measurements are the sides $\overline{OQ'}$ and $\overline{OR'}$. It is easy to see that the three angles of $\Delta OQQ'$ are as shown in the enlarged version on the right of figure 3.11. Then the sine law from elementary trigonometry tells us that

$$\frac{\overline{OQ'}}{\sin(90^\circ + \alpha)} = \frac{\overline{OQ}}{\sin(90^\circ - 2\alpha)}$$

and hence

$$\overline{OQ'} = \overline{OQ}\ \times \frac{\cos\alpha}{\cos(2\alpha)}$$

I leave it as an exercise for you to show that this can be rewritten as

$$\overline{OQ'} = (x - vt)\,\frac{\sqrt{1+\beta^2}}{1-\beta^2}$$

a very similar reasoning shows that

$$\overline{OR'} = (ct - \beta x)\,\frac{\sqrt{1+\beta^2}}{1-\beta^2}$$

We are early done - all that is left is to note that Bob's scale for units of length and time are larger than Alice's by a factor $\sqrt{\frac{1+\beta^2}{1-\beta^2}}$. Taking this into account it is easy to see that

$$\begin{aligned} x' &= \overline{OQ'} \div \sqrt{\frac{1+\beta^2}{1-\beta^2}} = \frac{x - vt}{\sqrt{1-\beta^2}} \\ ct' &= \overline{OR'} \div \sqrt{\frac{1+\beta^2}{1-\beta^2}} = \frac{ct - \beta x}{\sqrt{1-\beta^2}} \end{aligned}$$

which are, of course, the Lorentz transformations !

3.9 Causality and STR

Causality is perhaps the single most important concept of all physics - yet it is so simple to state that it is difficult to think of it as something non-trivial! So, just what is causality? It's the simple commonsense notion - *cause must occur before effect!* In a sense, causality is not a principle of physics at all - it is a part of what you can call proto-physics - something so fundamental that it takes precedence over the laws of physics. Any proposed law of physics that violates causality has to be rejected outright - that's how powerful this concept is!

In STR the issue of causality becomes especially important. Before I explain why, let me hasten to assure you that the concept is equally valid no matter whether you are talking of Newtonian or post-Newtonian physics. So, just why is it that people start talking about causality when it comes to STR? The answer lies in that feature which I have repeatedly called the most distinctive feature of STR - the transformation of time.

In Newtonian physics - the fact that cause must come before effect in time is a principle that has to be ultimately checked with experiments. However, if causality is valid for a particular cause-effect pair according to one observer, all other observers must agree that for this pair of events at least, causality works. This is because Newton's physics had one invariant absolute time for all - the time order of any two events (for that matter, even the size of the time interval), causally related or not, must be the same for all! So once causality is checked for one inertial observer, there is no further need to worry.

STR is different - simply because the time interval that Alice and Bob will measure for the same pair of events will differ - even to the extent of sign of the interval. Equation (3.25-d) shows that $\Delta t'$ can be negative, even though Δt is positive and *vice versa*. This means that while event $\mathcal{A}$ might have occurred prior to event $\mathcal{B}$ to Alice, it is possible for Bob to see $\mathcal{A}$ occur after $\mathcal{B}$!

Now if Alice sees $\mathcal{A}$ cause $\mathcal{B}$, then causality worked fine for her. What about Bob? He has no option but to agree that $\mathcal{A}$ causes $\mathcal{B}$, but for him, the order of the two events has been reversed! So, you see, in STR, the concept of causality is not very straightforward - in particular, we can see that there is a very real threat to its validity! This means that we have to take a much more careful look at the concept of causality when we apply it to STR.

Let's see how fast Bob must move so that a time interval that if positive to Alice becomes negative for him. Equation (3.25-d) tells us that this will happen if

$$\Delta t - \frac{v\Delta x}{c^2} < 0$$

which means that

$$v > \frac{c^2\Delta t}{\Delta x} \tag{3.46}$$

Now, if the event $\mathcal{A}$ has caused event $\mathcal{B}$, then there must have been some sort of influence that travels from the location of the former to that of the latter. At the very least, the "news" that $\mathcal{A}$ has occurred has to reach the location of $\mathcal{B}$, on or before it occurs. Thus, from Alice's point of view, the news has to cover the distance Δx in at most a time Δt. Thus the speed at which the information travels must satisfy

$$v_{\text{info}} \geq \frac{\Delta x}{\Delta t}$$

Thus, if $\mathcal{A}$ is to cause $\mathcal{B}$ *and* Bob is to see $\mathcal{B}$ occur before $\mathcal{A}$, we must have

$$v_{\text{info}} v > c^2$$

which means that at least one of the two speeds $v_{\text{info}}$and v must exceed c!

This, in fact is the strongest reason why nothing can travel faster than light. If the information that $\mathcal{A}$ has occurred could get to the location of the event $\mathcal{B}$ faster than light, then it is possible that Bob will be in a position to see the effect occur before the cause, even if his own speed is within the light-speed limit! On the other hand, if Bob's less-than-light speed satisfies the inequality (3.46), then Bob will say that $\mathcal{B}$ has preceded $\mathcal{A}$ while Alice claims just the opposite. However, in order that either of the two events cause the other one, it would be necessary for the information that the cause has occurred to reach the effect's site faster than light. Banning the possibility of faster than light propagation of the signal saves us from the prospect, then, that one of the two friends will see the effect occur before the cause!

To round off the discussion, let me stress that although STR does allow the possibility that the time order between two events will be different for two different observers, as long as both signals and observers obey the light speed limit, there is no conflict with the principle of causality. In the jargon of STR, we say that such reversal of time order for two observers is possible only for two events that are not causally connected.

In order to avoid any possible misunderstanding, let me emphasize that if event $\mathcal{B}$ is within the causal future of event $\mathcal{A}$, then it is *possible* that $\mathcal{A}$ is the cause of (or one of the causes of) event $\mathcal{B}$. This does not mean that $\mathcal{A}$ *has* to be the cause of $\mathcal{B}$, however! Of course, $\mathcal{A}$ can not be regarded as the cause of the infinite number of other events that fall inside its future lightcone. On the other hand, if $\mathcal{B}$ is not within the causal future of $\mathcal{A}$, then it is impossible for $\mathcal{A}$ to cause $\mathcal{B}$. A similar set of statements could be made for the case where $\mathcal{B}$ lies in the past lightcone of $\mathcal{A}$.

The splitting up of spacetime intervals into timelike, spacelike and lightlike picks up special significance in light of the discussion above. Rewriting (3.46) slightly shows that two observers will disagree about the relative time order of two events only if

$$\frac{v}{c}\Delta x > c\Delta t$$

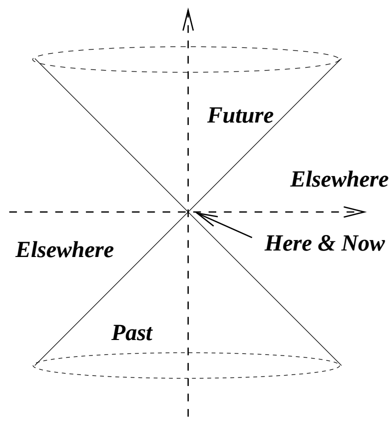
and since $\frac{v}{c} < 1$, this requires

$$\Delta x > c\Delta t$$

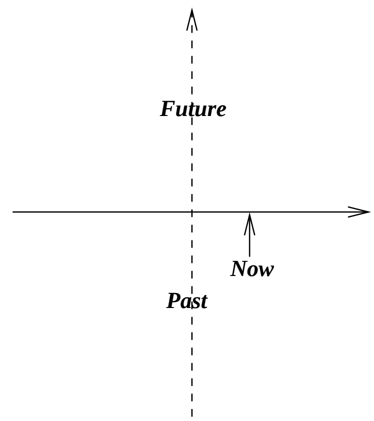
So, temporal order can be reversed only if the two events are separated by a spacelike interval. On the other hand - this means that the distance at which the two events occur exceeds c times the time interval, *i.e.* the distance is more than that even light can cover in this time. Thus, two events separated by a spacelike interval cannot be linked by a cause-effect relation.

On the other hand, if the interval between a pair of events is timelike or lightlike, then the time interval is sufficiently large for a slower-than-light (or light) signal to cover the distance between them. In this case, it is possible that the earlier event was the cause of the latter one. Note that in this case, all observers (as long as they stay slower than light) will agree about the temporal order in which the two events occur - so as long as causality holds for one observer, it will hold for all of them.

The causal structure of spacetime in relativity can be depicted very forcefully by means of the lightcone that I showed you earlier. Figure 3.12 shows the lightcone centered at a particular event - labelled *"Here & Now"*. The lightcone comes in two halves - for obvious reasons the part for positive time is called the *future light-*



(a) STR



(b) Newtonian

Figure 3.12: The causal structure of spacetime

cone, while the other half is labelled the *past lightcone.* The future lightcone and its interior contains points which are separated from the origin by a timelike (or lightlike) interval, and which, moreover, lies to its future - this is the *causal future* of the origin. The event at the origin can be the cause of events in this spacetime region. Similarly, the past lightcone and its interior constitutes the *causal past* of the origin. The events here can be the cause of the events at the origin. The region of spacetime lying outside the lightcone consists of events that are separated by a spacelike interval from the event at the origin. The appropriate label that we can assign such events is *elsewhere* - no observer will ever see any of them occur at the same place as the one at the origin.

The figure also shows two dashed lines - the X axis and the T axis. Why have I used dashed lines for these? Well, these lines are relevant only for a particular observer - as I have shown you earlier, the corresponding axes will be quite different for any other observer moving with respect to this one. Note that events that are in the causal future of the origin occur later than the origin, for all observers (we are sticking to observers that do not exceed the speed of light) - so the “future” label assigned to this region is not specific to a particular observer. The same goes for the causal past. As we have seen already, events that are separated from the origin by a spacelike interval have no specific time order - an event in this region may occur before the origin for one observer, but may very well occur after it for another observer. This is precisely why we don’t use the words past or future in the context of these events - the only observer independent designation that they may be given is the one that we have given them - elsewhere!

So that you can appreciate how different things where in the pre-Einstein world, I have also drawn the Newtonian view of the causal structure of spacetime in figure 3.12. In that diagram you will notice that I have drawn the X axis with a solid line, unlike the diagram for STR. This is because this line consists of all events that are simultaneous with the event at the origin - and this designation is something that every observer agrees with (this is Newtonian physics - remember?). This line, labelled *now* separates spacetime into two halves - the past and the future. Note that since Newtonian physics places no upper limit on the speed of signal propagation, the event at the origin could have been caused by any event in its past, while it could be the cause of any event in its future. I am running the risk of being subjective here - but I do feel that the causal structure of STR is much more beautiful than the rather dull setup that Newtonian physics provides!

Chapter 4

Light - the messenger of relativity!

4.1 The Doppler effect, again

Since the Doppler factor plays such an important role in our story, it is only fitting that we devote a bit more time to the effect from which it derives its name. As you all know, the Doppler effect is what makes a train's whistle sound high-pitched when it is approaching you and lowers the pitch when it is moving away. All of us have noticed the sharp drop in the pitch of the train whistle as a galloping train rushes by us. The corresponding effect in light has some subtleties due to relativity as we have seen. In this section, I am going to describe these in some detail.

4.1.1 The longitudinal Doppler effect

We have already seen this effect in action. This, precisely is what gives the Doppler factor its name. Bob rushes away from Alice, who shines a flashlight at him. While Alice sees each successive crest of the electric field in light at a time interval of τ_0, Bob sees them at a larger time interval. There are two factors at work here, as I have already shown you. Firstly, the crests have to catch up with Bob as he rushes away. This causes the lowering of the frequency that is expected even in classical physics. If this had been the only effect, the frequencies as measured by Bob and Alice would have been related by

$$\nu_{\text{Bob}} = \nu_{\text{Alice}} (1 - \beta) \tag{4.1}$$

This precisely is the same as the formula for the Doppler effect in sound, too - for the observer receding from a source of sound stationary in air. Of course, for sound, the ratio β will have to be changed to $\frac{v}{v_{\text{S}}}$, where v_{S} is the speed of sound. Secondly, Bob's clock runs slow, making the time interval between the crests smaller than that predicted by the first effect. This makes the actual frequency that Bob sees larger than that predicted classically (but still smaller than what Alice measures). Taking both factors into account, the frequency that Bob actually sees is

$$\nu_{\text{Bob}} = \nu_{\text{Alice}} \frac{1-\beta}{\sqrt{1-\beta^2}} = \nu_{\text{Alice}} \sqrt{\frac{1-\beta}{1+\beta}} \tag{4.2}$$

The Doppler effect gives us a simple way of checking the predictions of STR. All one has to do, it seems, is measure the frequency of light emitted by a source that is moving with respect to me, and verify whether the shift in frequency obeys the classical law, (4.1), or the relativistic one, (4.2).

Sources of light that are moving are very common, indeed! Even the light bulb that fixed on the wall provides such sources, even if you are sitting down right in front of it! You may have caught on to what I am hitting at here from the fact that I used the plural -

each individual atom that radiates the light photons are in violent to and fro motion. Atomic speeds are quite rapid, of the order of the speed of sound or more. This might seem like a small fraction of the speed of light - but remember that optical measurements like that of frequency can be carried out to very high precision.

There is one small hitch, though. In this case, the classical effect is already quite large - what STR does is that it introduces a small correction to this result (at least, the correction is small when the source is not moving very fast). To see this, let me write the expression for the frequency as seen by Bob as a series expansion in β, which is almost always rather small, to get

$$\nu_{\text{Bob}} = \nu_{\text{Alice}} \left(1 - \beta + \frac{1}{2}\beta^2 + \dots\right) \tag{4.3}$$

Unless β is rather large, the classical effect $-\beta$ swamps out the STR term of $\frac{1}{2}\beta^2$ - and hence the relativistic effect that time dilation causes can be rather difficult to detect directly.

In spite of what I have said right now, experimentalists have been clever enough to actually detect the small relativistic effects. They were helped by the fact that the classical effect depends linearly on β - which means that the frequency shifts by opposite amounts due to it for atoms moving in opposite directions. The atoms move randomly due to thermal effects - which means that the average frequency stays the same due to the classical term. This term has a big effect, though - it causes the sharp lines of an atomic spectrum to broaden out - something that spectroscopists call Doppler broadening. Any shift in the average frequency, though, must be caused by the relativistic term(s). Despite the extreme smallness of this shift, experimentalists have managed to use a wonderful effect of solid state physics called the Mossbauer effect to detect this - providing a direct verification of relativity.

4.1.2 The transverse Doppler effect

Although we have been sticking to a one dimensional world for simplicity, this is as good a point as any in which to introduce the other directions as well. What happens if Alice and Bob were to both receive a light signal when they were side by side, coming from a source that is standing still with respect to Alice but displaced from her in a direction that is perpendicular to Bob's speed? For such transverse beams of light, Bob is not really running away from the source. Since light has no catching up to do, classical physics says that both Alice and Bob will receive the electric field maxima at the same time - and hence will see the same frequency!

There is a small catch here. Alice will keep on receiving the successive peaks, one by one. Bob, however, will catch only one (the one that reaches them both when they are side by side) directly. When the next one reaches Alice, Bob has already moved slightly (unless he is moving really fast!) to one side. Thus Bob has to time the arrival of the second pulse at Alice's location indirectly! So, while Alice measures the proper time between the pulses, Bob does not! Since Bob concludes that Alice's watch is running slow, he must have measured a bigger time gap and hence a smaller frequency for the light, purely due to the time dilation effect! It is very easy to figure out the relation between the two frequencies

$$\nu_{\text{Bob}} = \nu_{\text{Alice}}\sqrt{1-\beta^2}$$

where v is Bob' speed with respect to Alice (as well as the source of light). Note that Alice receives the peaks at exactly the same interval as the source of light itself, so the frequency she measures is the proper frequency, while Bob measures a smaller one. This transverse Doppler effect, then, is a purely relativistic effect - occurring entirely due to time dilation.

You may be tempted to argue that Alice must see Bob's clock run slow too, and it is she who should be measuring the lower frequency! Remember, though, that Alice receives all electric field maxima at the same place, and hence is the one measuring the proper time between them. The role of the two friends isn't really symmetrical here, it is Alice who is stationary with respect to the source of light, and Bob who is moving with respect to it!

4.2 More on light

4.2.1 Aberration

4.2.2 Propagation

Chapter 5

Additional topics in kinematics

In this chapter I am going to dwell on several important physical consequences of relativistic kinematics that we have derived above Although they are very important, if you are in a hurry, you can skip ahead to the next chapter, where I deal with the laws of mechanics in relativity. I would of course prefer that you read all that follows (after all, that is why I am writing all this!)! Even if you decide to skip ahead, do come back later to these topics.

5.1 Deriving the Lorentz transformations, directly

5.2 A new notation

As I will explain in some more detail a bit later, in a sense relativity places time on an equal footing with space. It is easier to see this if you measure both space and time in the same units (so that, as I mentioned earlier, the speed of light comes out to be unity). If you want to stick to the same units that you have been using all along, the same effect can be obtained by considering the product ct instead of time.

I am going to use the abbreviation x^0 for ct, while I will use the obvious nomenclature $x^1 = x$, $x^2 = y$ and $x^3 = z$ and a similar notation for the spacetime coordinates according to the primed observer. As you can see, I have chosen to write the indices on the coordinates as superscripts rather than the usual subscripts. Why? There is a very good reason - one that will hopefully become clear in a while from now.For the time being, do pay attention to the fact that x^2refers to the second component of the quantity x and not the square of x. If I mean the latter, I will write it explicitly as $(x)^2$. In terms of these,the Lorentz boost equations in (3.23-a-3.23-d) becomes

$$x'^0 = \gamma x^0 - \beta\gamma x^1 \quad (5.1\text{-a}) \qquad x^0 = \gamma x'^0 + \beta\gamma x'^1 \quad (5.2\text{-a})$$
$$x'^1 = -\beta\gamma x^0 + \gamma x^1 \quad (5.1\text{-b}) \qquad x^1 = \beta\gamma x'^0 + \gamma x'^1 \quad (5.2\text{-b})$$
$$x'^2 = x^2 \quad (5.1\text{-c}) \qquad x^2 = x'^2 \quad (5.2\text{-c})$$
$$x'^3 = x^3 \quad (5.1\text{-d}) \qquad x^3 = x'^3 \quad (5.2\text{-d})$$

This can obviously be written as the matrix equation

$$x' = Lx \qquad x = L^{-1}x'$$

where x and x' are the 4×1 column matrices

$$x = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}, \qquad x' = \begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix}$$

while L is the 4×4matrix

$$L = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

If you are afraid of matrices, you can also write this in terms of components

$$x'^{\mu} = \sum_{\nu=0,1,2,3} L^{\mu}{}_{\nu} x^{\nu} \tag{5.3}$$

where obviously L^{μ}_{ν} is the entry at the μ-th row and ν-th column of the matrix L. You are certainly familiar with this entry being denoted $L_{\mu\nu}$ - so why this perverse notation - in which the row index is written as a superscript and the column index as a subscript? I will explain this in due time - for the time being please bear with this!

I have written down the coefficients explicitly for the special Lorentz boost. It should be obvious that the compact form of the transformation equations, (5.3) is valid for the general Lorentz transformation that I discussed in the last subsection. Of course, the individual coefficients may be much more complicated than in our special case. A natural question that will arise is, given a set of four equations that look like (5.3), can we be sure that they stand for a Lorentz transformation? You will have to wait a while for the answer - which I will provide in the next section.

5.3 General Lorentz transformations and the invariance of the interval

I have already shown you that the interval has to stay the same under a general Lorentz transformation. This has some immediate consequences for the coefficients of a general Lorentz transformation themselves.

Remember that we can write a general Lorentz transformation as

$$\text{either} \qquad x' = Lx, \qquad \text{or} \qquad x'^{\mu} = \sum_{\nu=0,1,2,3} L^{\mu}_{\nu} x^{\nu}$$

where on the left I have written the matrix form and on the right, the explicit form in terms of the transformation coefficients (subsection 5.2).

5.3.1 The covariant coordinates

Now, in this notation, it is easy to see that I can write the invariant interval as

$$\Delta s^2 = \left(\Delta x^0\right)^2 - \left(\Delta x^1\right)^2 - \left(\Delta x^2\right)^2 - \left(\Delta x^3\right)^2$$

If all the signs above had been +, it would be child's play to write the interval in a compact form - $\Delta s^2 = \sum_{\mu=0}^{3} (\Delta x^{\mu})^2$. In terms of the matrix x, this would have been even more compact[1], $\Delta s^2 = (\Delta x)^T (\Delta x)$! Of course, all the signs are not the same - so this simple form is not the right one. However, it is very easy to see that if you were to introduce another set of four numbers to coordinatize

[1] Here, the superscript T stands for the transpose of a matrix - the matrix that is obtains by it by interchanging rows and columns.

spacetime,

$$x_0 \equiv x^0 = ct \tag{5.4-a}$$
$$x_1 \equiv -x^1 = -x \tag{5.4-b}$$
$$x_2 \equiv -x^2 = -y \tag{5.4-c}$$
$$x_3 \equiv -x^3 = -z \tag{5.4-d}$$

and use *both* sets of coordinates together, a compact notation can be achieved.

$$\Delta s^2 = \sum_{\mu=0}^{3} (\Delta x^\mu)(\Delta x_\mu) \tag{5.5}$$

If I define a column vector $\underline{x}$, as

$$\underline{x} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

then I can also write a nice and simple matrix form

$$\Delta s^2 = (\Delta \underline{x})^T (\Delta x). \tag{5.6}$$

A small bit of nomenclature - the set of numbers (x^0, x^1, x^2, x^3) is called the contravariant position vector in spacetime - while (x_0, x_1, x_2, x_3) is called the covariant position vector. This terminology will become clearer once we study vectors in spacetime in section 5.5.

Since we know how (x^0, x^1, x^2, x^3) transform under a Lorentz transformation, it is easy to figure out the way in which the covariant coordinates change. In the case of the special Lorentz transformation

$$\begin{aligned} x'_0 &= \gamma x_0 + \beta\gamma x_1 && \text{(5.7-a)} \\ x'_1 &= \beta\gamma x_0 + \gamma x_1 && \text{(5.7-b)} \\ x'_2 &= x_2 && \text{(5.7-c)} \\ x'_3 &= x_3 && \text{(5.7-d)} \end{aligned} \qquad \begin{aligned} x_0 &= \gamma x'_0 - \beta\gamma x'_1 && \text{(5.8-a)} \\ x_1 &= -\beta\gamma x'_0 + \gamma x'_1 && \text{(5.8-b)} \\ x_2 &= x'_2 && \text{(5.8-c)} \\ x_3 &= x'_3 && \text{(5.8-d)} \end{aligned}$$

In terms of the components, the transformation of the covariant coordinates are written

$$x'_\mu = \sum_{\nu=0}^{3} L_\mu^{\ \nu} x_\nu$$

The notation, $L_\mu^{\ \nu}$ is standard. As you must have guessed, it is the μ-th row and ν-th column element of a matrix, just as $L^\mu_{\ \nu}$ was. what might be some source of confusion is that despite the same letter L being used, they are not elements of the same matrix - indeed, it is the position of the indices that tell you which matrix we are talking about. As I had already said before, $L^\mu_{\ \nu}$ is the $\mu-\nu$-th element of the Lorentz transformation matrix L. One look at (5.2-a-5.2-d) and (5.7-a-5.7-d) might make you think that the other set of coefficients $L_\mu^{\ \nu}$ are really elements of the matrix L^{-1}. Actually they are elements of the matrix $\left(L^T\right)^{-1}$ - it is just that the matrix L happens to be symmetric for the special Lorentz transformation! How can you tell? Well, you will have to wait a little while for the answer!

I have just now answered a question that must have been bothering you since I introduced the new notation in subsection 5.2 - why the superscripts? Of course, I had to use superscripts there, because I have another set of four numbers that I also want to denote by the letter x. Because I was going to use the subscripts here, I had to use the superscripts there!

All you do to get the covariant position vector from the contravariant one is - simply change the sign of the spatial vector - so while the latter is $(ct, \vec{r})$, the former is $(ct, -\vec{r})$. The difference is so slight that at this stage it is natural to feel slightly cheated - this looks like a mere sleight of hand, a change of notation simply so that I can write a formula slightly more simply! At this stage, that is just what this is! You might even feel that having to deal with indices above and below is too much of a price to pay for this simplicity of notation. Have faith - this new addition to our notational stable is going to pay off rich dividends later.

Let's look back at just what has been going on here. Why did I have to bring in this new set of coordinates with subscripts on them? Because the interval, which the Lorentz transformation leaves invariant, had its signs all mixed up. Put a bit formally, this means that the Lorentz transformations are not orthogonal - they do not preserve the sum of squares of coordinates (which is the square of the length, according to good old Pythagoras). This forces us to bring in this new kind of entity. Of course, since the Lorentz transformations are only very slightly non-orthogonal (what they do preserve is very nearly the sum of squares - except for a few wayward signs!), the difference between the two versions of coordinates is very slight. If you had to deal with transformations that are even more badly non-orthogonal (and you would have to, if you wanted to study Einstein's General Theory of Relativity!), the difference between the contravariant and the covariant version of the coordinates would have been even more marked.

5.3.2 The metric

I have already given you a prescription of how to change a contravariant vector into a covariant one - all you have to do is change the sign of the space coordinates, keeping the time part intact! This can, of course be written in the form of a matrix η acting on the column vector $(x^0, x^1, x^2, x^3)^T$, where the matrix η, called the metric, is a diagonal matrix diag $(1, -1, -1, -1)$:

$$\underline{\mathbf{x}} = \eta x \tag{5.9}$$

or, more explicitly,

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{5.10}$$

or, in terms of components,

$$x_\mu = \sum_{\nu=0}^{3} \eta_{\mu\nu} x^\nu$$

where, of course

$$\eta_{00} = -\eta_{11} = -\eta_{22} = -\eta_{33} = 1$$

are the only nonzero elements of the matrix η.

We have seen the metric in its role as a converter from the contravariant coordinates to the covariant coordinates. Another way to look at this will become apparent if you take a look at the way the invariant interval is defined

$$\begin{aligned} \Delta s^2 &= \sum_{\mu=0}^{3} (\Delta x^\mu)(\Delta x_\mu) \\ &= \sum_{\mu=0}^{3} (\Delta x^\mu) \left(\sum_{\nu=0}^{3} \eta_{\mu\nu} \Delta x^\nu \right) \end{aligned}$$

$$= \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} \Delta x^{\mu} \Delta x^{\nu} \tag{5.11}$$

which you can interpret as the formula that will give you the "length" of a spacetime coordinate difference - in the same way as the length of a vector in three dimensions is

$$\sum_{\mu=1}^{3} \Delta x^{\mu} \Delta x^{\mu}$$

This tells us the reason behind the name metric - it is what is helpful in measuring "lengths" in spacetime. In terms of the matrix η, you can write the interval as

$$\Delta s^2 = (\Delta x)^T \, \eta \, (\Delta x) \tag{5.12}$$

The fact that the interval is invariant under a Lorentz transformation tells us that the Lorentz transformation matrix L has to obey a very important restriction. Since $x' = Lx$ obeys

$$(\Delta x')^T \, \eta \, (\Delta x') = (\Delta x)^T \, \eta \, (\Delta x)$$

we have

$$(\Delta x)^T \, L^T \eta \, L \, (\Delta x) = (\Delta x)^T \, \eta \, (\Delta x)$$

for all values of (Δx). This shows[2] that the matrix L has to obey

$$L^T \eta L = \eta \tag{5.13}$$

which is the condition that can tell us whether a particular 4×4 transformation matrix can qualify as a Lorentz transformation or not. In terms of the coefficients it is easy to see that this condition translates to

$$\sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} L^{\mu}{}_{\rho} L^{\nu}{}_{\sigma} = \eta_{\rho\sigma} \tag{5.14}$$

I told you a while ago that the covariant version of spacetime coordinates transforms according to the rule $\underline{\mathbf{x}}' = \underline{\mathbf{L}}\,\underline{\mathbf{x}}$, where the transformation matrix $\underline{\mathbf{L}}$ is related to the Lorentz transformation matrix L by

$$\underline{\mathbf{L}} = \left(L^T\right)^{-1}$$

This is easy to prove now. Note that

$$\underline{\mathbf{x}}' = \eta x' = \eta L x = \eta L \eta^{-1} \underline{\mathbf{x}}$$

so that the transformation matrix for covariant coordinates is $\underline{\mathbf{L}} = \eta L \eta^{-1}$. From equation (5.13) it follows that $\underline{\mathbf{L}} = \left(L^T\right)^{-1}$, as promised!

An important property of any Lorentz transformation matrix can be derived by taking the determinant of both sides of equation

[2]Actually, the fact that $x^T A x = x^T B x$ for all values of x does not really force us to conclude that $A = B$. All this really says is that the matrix $C = A - B$ obeys $x^T C x = 0$ for every column vector x. as you can easily check, this is satisfied, not only by the obvious choice - the null matrix, but also by a (and only by) any antisymmetric matrix. however, in this case it is easy to see that $L^T \eta L - \eta$ is a symmetric matrix - and is thus both symmetric and antisymmetric. The only matrix that is both *is* the null matrix.

(5.13), which immediately gives

$$(\det L)^2 = 1$$

and thus

$$\det L = \pm 1 \tag{5.15}$$

Thus all Lorentz transformations fall into two categories - one with $\det L = +1$ and the other with $\det L = -1$.

Another property of the Lorentz transformation matrix is easier to derive from the condition on coefficients, equation (5.14). Writing the equation for $\rho = \sigma = 0$ leads to

$$\sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} L^{\mu}{}_{0} L^{\nu}{}_{0} = \eta_{00} = 1$$

which can be rewritten in the form

$$\left(L^{0}{}_{0}\right)^2 - \sum_{i=1}^{3} \left(L^{i}{}_{0}\right)^2 = 1$$

leading to

$$\left(L^{0}{}_{0}\right)^2 = 1 + \sum_{i=1}^{3} \left(L^{i}{}_{0}\right)^2 \geq 1$$

and thus we have

$$\text{either} \qquad L^{0}{}_{0} \geq +1, \qquad \text{or} \qquad L^{0}{}_{0} \leq -1 \tag{5.16}$$

Putting the conditions (5.15) and (5.16) together we see that all Lorentz transformations can be grouped under four categories

a) $L_{+}^{\uparrow}$ Lorentz transformations with $\det L = +1$ and $L^{0}{}_{0} \geq +1$.

b) $L_{-}^{\uparrow}$ Lorentz transformations with $\det L = -1$ and $L^{0}{}_{0} \geq +1$.

c) $L_+^\downarrow$ Lorentz transformations with $\det L = +1$ and $L^0{}_0 \leq -1$.

d) $L_-^\downarrow$ Lorentz transformations with $\det L = -1$ and $L^0{}_0 \leq -1$.

The special Lorentz transformation obviously falls in the first category - which is called, rather pompously, the set of proper orthochronous Lorentz transformations.

5.4 The geometry of special relativity

Even if you had not learned any relativity before this, you are sure to have heard of one of its results - that STR tells us that we live in a four dimensional world - and that time is the fourth dimension[3]. In this section we will take a slightly detailed view of what this really means.

At the basic mathematical level, there is nothing surprising in the fourth dimension - all that we are saying is that you need four numbers to characterise an event - the three space coordinates that tell you where the event has occurred and one more number specifying the time when it occurred. This is true, of course, even in Newtonian physics - so just what's the big deal in relativity?

The big deal is, as we have seen over and over again already, there is a big difference in the way in which time behaves in Newtonian and relativistic physics. In Newtonian physics, time does enter into the equations for Galilean transformation - but more as a passive parameter. According to these transformations, all observers measure the same value for the time - it is absolute! In the Lorentz transformations, on the other hand, time is no longer in-

[3]If the idea of a fourth dimension above the usual three makes you feel a bit uncomfortable - brace yourself. Superstring theory will have us believe that we live in a ten dimensional spacetime!

variant - it mixes with space in much the same way the coordinates do when you rotate the coordinate axes.

Before I explain what is meant precisely by the statement that spacetime is four dimensional, let me take you back into the depths of a statement that you must be much more comfortable with - the statement that the space that we see around us is three dimensional. Of course, every schoolboy knows that you need three numbers to locate a point in space. However, we are also used, especially in dynamics problems, to treating the various space coordinates independently - you write down separate equations for the X component, the Y component and so on. Why is it, then, that we do not regard the three components as three *distinct* quantities, but rather look upon them as manifestations of one basic object, three faces of the same coin - so to speak?

The answer is obvious - which component is X, which one is Y ... are, after all, artificial choices. I could just as easily have chosen to orient my coordinate axes in a different way - this would mix all the previous components together to give the new component. It is this mixing up that shows that we have to regard the three spatial coordinates as a single entity, unless we force ourselves to keep on using one particular coordinate system. The latter would have been a reasonable option, though, if one coordinate system could have been somehow different from all others - which is not the case!

In much the same way, I could have kept on regarding time as completely separate from space if I had one inertial observer whom I could regard as special. However, the whole point of relativity is that this can not be done - all inertial observers are created equal! Add to this the fact that changing over from one inertial observer to another mixes time up with the space coordinates as well as *vice versa* and you have to regard the time along with the space coordinates as four pieces of the same thing that we call, for want of a better name, spacetime.

5.4.1 Coordinate changes in rotations

I have used a comparison between rotations and Lorentz transformations to explain why spacetime should be regarded as a four dimensional identity with x, y, z and t as its components. The connection between rotations and Lorentz transformations run quiet a lot deeper than you may think. In this subsection I will take a deeper look at the mathematics of rotations to lay the groundwork for its comparison with the Lorentz transformations in the next subsection. At the very outset, we must take a look at the way in which the x, y, z coordinates transform when the coordinate system is rotated around. This mathematics will also come in handy when we discuss vectors and scalars in three and four dimensions in section 5.5.

To keep things simple I will only consider a rotation through an angle ϑ about a single axis - let's say the Z axis, as shown in figure 5.1. Of course, the z coordinate stays intact under such a rotation. What about the other two coordinates? It is clear from the figure that

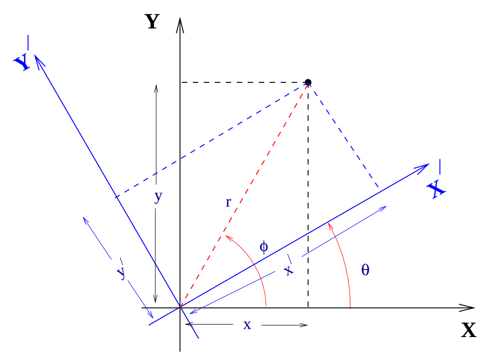


Figure 5.1: Rotating the coordinate system

$$x = r\,\cos\phi \qquad x' = r\,\cos(\phi - \vartheta)$$
$$y = r\,\sin\phi \qquad y' = r\,\sin(\phi - \vartheta)$$

from which a bit of the trigonometry that you learnt in high school will take you to

$$x' = x\cos\vartheta - y\sin\vartheta \tag{5.17-a}$$

$$y' = x\sin\vartheta + y\cos\vartheta \tag{5.17-b}$$

$$z' = z \tag{5.17-c}$$

If you are familiar with matrices, you will of course realize that this can be written in the compact matrix form

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

where the matrix

$$R_z(\vartheta) = \begin{pmatrix} \cos\vartheta & -\sin\vartheta & 0 \\ \sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is called the rotation matrix for a rotation through an angle ϑ about the Z axis. You should realize that the rather simple form that this matrix has is due to the fact that I have considered a rather special rotation - the matrix for a rotation about an arbitrary axis looks much more complicated (no columns or rows made out of 0's and 1's!) - but is actually just as simple in concept. If you haven't learned about matrices yet, don't worry - we will not use them any further here.

5.4.2 Lorentz transformations and rotations

As in the case of rotations, we first start with a look at a particular Lorentz transformation - the boost along the X axis.

One thing should be clear from (5.17-a-5.17-c) and (5.1-a-5.1-d) - which is the rather deep similarity between rotations and the Lorentz boost. Of course, there are some differences, too. The most obvious of them is, of course, that in the Lorentz transformations t plays a very active role - while it is conspicuous by its absence in rotations. A direct result of this is that while rotations involve only three variables - the Lorentz transformations involve four. From a purely mathematical point of view, this is not a big deal, though! What makes the real difference between the rotations and Lorentz transformations is what quantities they keep invariant. While for rotations the invariant quantity is the length

$$l^2 = x^2 + y^2 + z^2$$

for Lorentz transformations the invariant quantity is the interval

$$c^2t^2 - x^2 - y^2 - z^2.$$

This difference is compactly expressed by what is called the *signature* - while this is $(+,+,+)$ for rotations, it is $(+,-,-,-)$ for the Lorentz transformations. This makes the Lorentz transformations non-orthogonal (orthogonal transformations, remember, preserve the *sum* of squares of the coordinates), but not by very much!

This non-orthogonality of the Lorentz transformations reveals itself in the coefficients of our special Lorentz transformations, too. Note that the two coefficients in the equation for x'^0 are γ and $-\beta\gamma$, whereas the corresponding quantities for the rotation are $\cos\vartheta$ and $\sin\vartheta$. As you have known for ages now, the latter coefficients obey

$(\cos\vartheta)^2 + (\sin\vartheta)^2 = 1$. what about the Lorentz transformation coefficients? The sum of their squares is not something simple, but their difference is $(\gamma)^2 - (-\beta\gamma)^2 = (1-\beta^2)\gamma^2 = 1$! I leave it to you to figure out just what the connection is between the orthogonality issue and this.

Let me point out, though, that we do know of simple functions that can take the place of the $\cos$and sin in the Lorentz transformations. The fact that the *difference* of squares of the coefficients is unity, immediately suggests using the hyperbolic functions - $\cosh\phi = \gamma$ and $\sinh\phi = \beta\gamma$. The parameter ϕ is easily seen to be

$$\phi = \tanh^{-1}\beta$$

The parameter $\phi$is rather similar to our Doppler factor K in the sense that it too provides an alternative to the speed as a means of parameterising how fast a body is moving. I personally prefer the doppler factor, especially since it is easily measurable experimentally. However, the use of ϕ rather than the speed does simplify some calculations - so it is important to remember this alternative. This parameter is given a special name - it is called the **rapidity**.

5.4.3 A bit of history - Minkowski coordinates

You should be able to see a very simple trick which will make the Lorentz transformations orthogonal. All you have to do is to ensure that is preserves the sum of squares of the coordinates by taking as the fourth coordinate as, not ct, but ict where i is the familiar square root of minus 1! In order to distinguish between our earlier convention, I will call this fourth coordinate x_4 instead of the x^0 that we were using earlier. In addition, I will use subscripts instead of superscripts on the three space coordinates as well - $x_1 = x$, $x_2 = y$

and $x_3 = z$[4]! The person who introduced these coordinates into STR is Einstein's friend, the mathematician Hermann Minkowski. In his honour, this choice of coordinates is called the Minkowski coordinates.

In terms of these coordinates the invariant quantity is, of course

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 \equiv x^2 + y^2 + z^2 - c^2t^2 \tag{5.18}$$

and the special Lorentz transformations take the form

$$x'_1 = \gamma x_1 + i\beta\gamma x_4 \quad (5.19\text{-a}) \qquad x_1 = \gamma x'_1 - i\beta\gamma x'_4 \quad (5.20\text{-a})$$
$$x'_2 = x_2 \quad (5.19\text{-b}) \qquad x_2 = x'_2 \quad (5.20\text{-b})$$
$$x'_3 = x_3 \quad (5.19\text{-c}) \qquad x_3 = x'_3 \quad (5.20\text{-c})$$
$$x'_4 = -i\beta\gamma x_1 + \gamma x_4 \quad (5.19\text{-d}) \qquad x_4 = i\beta\gamma x'_1 + \gamma x'_4 \quad (5.20\text{-d})$$

The similarity between these equations and those for rotation should be obvious. Indeed, even the sum of squares of the two coefficients γ and $(-i\beta\gamma)$ is 1, so that you can even identify them with $\cos\vartheta$ and $\sin\vartheta$, respectively! Of course, in this case we will get $\tan\vartheta = -i\beta$ so that the angle involved is complex (That's obvious, since $\cos\vartheta = \gamma \geq 1$!). Thus, in the Minkowski coordinates Lorentz transformations are seen to be rotations, but in the $x_1 - x_4$plane, in the same way as the rotation about the Z axis was one in the $x_1 - x_2$ plane.

This identification of the Lorentz transformations with rotations was one reason which made Minkowski coordinates very popular in their day. However, the fact that the rotation angle had to be complex actually makes this identification rather tricky! The

[4]"There he goes again!" - I can hear you groaning! Why do I revert to subscripts here? Why were I using superscripts is the first place? All that I can tell you now is, as before, keep patience! All will be revealed in a short while!

use of imaginary time made time look very much the same as the space coordinates - but this was an illusory similarity. In fact, they were instrumental in giving people the notion that relativity places "space and time on the same footing".

It is true that relativity does bring time a lot closer to space than it was thought earlier. The very fact that time ceases to be absolute, but rather observer dependent like space is a huge step in this direction. In spite of this there is a very big difference between space and time that persists in relativity - you can go in any direction in space - but in only one direction in time. Since causality tells us that cause always precedes effect in time, shows that this distinction is about as basic as can be! Indeed, hiding the different sign of t^2 in the invariant interval by using an imaginary time coordinate conceals this very important difference - which turns out to be more of a burden than an advantage in the long run. Thus, despite having some advantages, Minkowski coordinates are used only very rarely today!

5.4.4 Minkowski spacetime

Minkowski coordinates may have fallen out of favour today, but Minkowski's biggest contribution to STR has endured. This is the idea that one has to consider spacetime as a single four dimensional entity and what one observer calls space and time coordinates of an event are simply components of this entity in axes that are relevant for him or her. This is the Minkowski spacetime, in which points have the coordinates x^0, x^1, x^2, x^3 or, in more familiar terms $(ct, \vec{r})$. Let me stress once again that this clubbing together of space and time coordinates is not merely a matter of putting two different things in the same bracket - the Lorentz transformation mixes them up - so that the distinction between space and time is rather blurred in STR.

The geometry of Minkowski spacetime is in some ways very similar to that of our ordinary three dimensional space - but in some other ways it is very different. In ordinary space we have the Pythagoras theorem that gives the distance between two points as

$$d^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$$

- which stays invariant when you rotate your coordinate system. In Minkowski spacetime we have a new version of the Pythagoras theorem - the "distance" between two events in spacetime is given by the interval

$$\Delta s^2 = c^2 \Delta t^2 - \left(\Delta x^2 + \Delta y^2 + \Delta z^2\right)$$

which, of course stays unchanged when you move from one inertial observer to another.

As I have been pointing out over and over again, the similarity between the interval and the ordinary distance is what helps us to motivate the generalization of ordinary space to spacetime. At the same time, the difference between the interval with its $(+,-,-,-)$ signature and distance with its $(+,+,+)$ signature is substantial - and it is this difference that makes the geometry of spacetime very different from that of space. In particular, let me stress once again that despite its merger with space, time remains a very different entity from space even in STR so unlike space, where all three dimensions are equivalent - in spacetime we have a certain lack of democracy!

One very important distinction between space and spacetime is that the distance in four dimensions is *not* necessarily positive[5]!

[5]Indeed, if I try to push the analogy between distance and the interval, then I

The interval can be positive, zero or even negative! Of course, this is what gave us the distinction between spacelike, null and time-like intervals. Mathematicians call such a "distance" like function a *pseudo-metric*, in their language Minkowski space is a pseudo-metric space.

5.5 Four-vectors and four-scalars

When we were discussing the addition of velocities in STR in section 3.4, you must have been struck by the fact that the rather unusual quantities that I denoted by U_x, U_y and U_z transform much more simply from observer to observer than the familiar velocity components themselves. In this section I will show you why! But first, I need to take you through a short detour - to ensure that you understand what vectors and scalars really are.

5.5.1 Vectors and scalars in 3 dimensions

You may be understandably indignant at the last sentence - after all, every high school kid has learnt about vectors and scalars! However, to understand what four-vectors and four-scalars are, you will have to go a bit deeper into the meaning of ordinary vectors and scalars than what you have learnt in the high school.

To keep the discussion simple, I will take off from what you have always thought of vectors as being - arrows going from one point to another. Imagine a vector $\vec{A}$ stretching from the point P_1 to the point P_2. If the coordinates of the two points are (x_1, y_1, z_1) and (x_2, y_2, z_2), respectively, then, as you all know, our vectors will have

should talk of Δs rather than Δs^2. In that case, the four dimensional "distance" is not even always real!

components of

$$A_x = x_2 - x_1, \quad A_y = y_2 - y_1, \quad A_z = z_2 - z_1$$

Now, the coordinates of the points P_1 and P_2 certainly depend on the coordinate system that we are using - they will change when we rotate the coordinate system. It stands to reason, then, that the components of my vector $\vec{A}$ will change, too! Let's investigate what this change is like.

We have seen how the coordinates of a point changes under a rotation of the axes in equations (5.17-a - 5.17-c). From this it is child's play to figure out how the components of the vector $\vec{A}$ will change. Let me show you how to handle the X component

$$\begin{aligned} A'_x &= x'_2 - x'_1 \\ &= (x_2 \cos\vartheta - y_2 \sin\vartheta) - (x_1 \cos\vartheta - y_1 \sin\vartheta) \\ &= (x_2 - x_1)\cos\vartheta - (y_2 - y_1)\sin\vartheta \\ &= A_x \cos\vartheta - A_y \sin\vartheta \end{aligned}$$

You can find out A'_y and A'_z in a similar fashion. The result is a set of transformation equations

$$A'_x = A_x \cos\vartheta - A_y \sin\vartheta \tag{5.21-a}$$

$$A'_y = A_x \sin\vartheta + A_y \cos\vartheta \tag{5.21-b}$$

$$A'_z = A_z \tag{5.21-c}$$

If you take a look at the set of transformation equations (**??**-5.17-c) and (5.21-a-5.21-c) side by side you will immediately notice the striking fact that the set (A_x, A_y, A_z) of the components of a vector transform in exactly the same fashion as the coordinates (x, y, z). A moment's thought will show you that this will remain true even for more complex rotations, too.

There you have it - the definition of a vector that we will need for relativity. *An ordinary vector (formally known as a three vector) is a set of 3 numbers (called components) that transform when you rotate the coordinate system in exactly the same way in which the coordinates themselves do!* Another way to put this is - all vectors are 3 component objects that transform in the same way under coordinate rotations. What that transformation rule is is specified by the fact that the position vector is a prototype vector.

What is a scalar? Well, firstly it is something that is described by a single number. However, anything that is described by a single number does not qualify - it has to transform in a very special way! Indeed, the transformation law for a scalar quantity is so simple that it hardly seems to be a transformation law at all! A scalar simply does not change under a rotation of coordinates.

Let me now point out that one quantity that you had always thought of as a scalar - the length of a vector, is really a scalar[6]! To show this, let me start with the length of the prototype vector - the position vector, $l = \sqrt{x^2 + y^2 + z^2}$. That this is unchanged by rotations is no surprise - the fact that distances do not change is one of the basic defining features of rotations!

There is no real need for a proof that under a rotation length of a vector remains the same, $l = l'$, given that it follows from the definition of a rotation. I will verify that this works for the rotation for which I have written down the coordinate transformation in equations (5.17-a- 5.17-c) nevertheless! Here goes

$$l'^2 = x'^2 + y'^2 + z'^2$$

[6]If you are wondering what's the big deal with that - let me point out that some other quantities like the X component of a vector that you would be likely to think of as a scalar is not really a scalar according to this criterion!

$$
\begin{aligned}
&= (x\cos\vartheta + y\sin\vartheta)^2 + (-x\sin\vartheta + y\cos\vartheta)^2 + z^2 \\
&= x^2\cos^2\vartheta + 2xy\cos\vartheta\sin\vartheta + y^2\sin^2\vartheta + x^2\sin^2\vartheta \\
&\quad -2xy\cos\vartheta\sin\vartheta + y^2\cos^2\vartheta + z^2 \\
&= x^2\left(\cos^2\vartheta + \sin^2\vartheta\right) + y^2\left(\cos^2\vartheta + \sin^2\vartheta\right) + z^2 \\
&= x^2 + y^2 + z^2 = l^2
\end{aligned}
$$

- as promised.

Of course, what I have done here can be hardly considered a proof that rotations keep lengths unchanged. After all, all that has been done is a verification that this happens for a particular kind of rotation. However, since the choice of the Z axis is completely artificial - it can be argued that if lengths do not change for rotations about the Z axis, then they won't change for rotations about any axis.

Now that we have shown that the length of a position vector stays unchanged under rotations, what about the length of any other vector? Since all vectors transform in exactly the same way as the position vector, it is easy to see that this invariance of the length will work for all of them. If you are still sceptical, here is the detailed verification

$$
\begin{aligned}
A'^2 = \vec{A}' \cdot \vec{A}' &= A_x'^2 + A_y'^2 + A_z'^2 \\
&= (A_x\cos\vartheta + A_y\sin\vartheta)^2 + (-A_x\sin\vartheta + A_y\cos\vartheta)^2 + A_z^2 \\
&= A_x^2\cos^2\vartheta + 2A_xA_y\cos\vartheta\sin\vartheta + A_y^2\sin^2\vartheta + A_x^2\sin^2\vartheta \\
&\quad -2A_xA_y\cos\vartheta\sin\vartheta + A_y^2\cos^2\vartheta + A_z^2 \\
&= A_x^2\left(\cos^2\vartheta + \sin^2\vartheta\right) + A_y^2\left(\cos^2\vartheta + \sin^2\vartheta\right) + A_z^2 \\
&= A_x^2 + A_y^2 + A_z^2 = \vec{A}\cdot\vec{A} = A^2
\end{aligned}
$$

Note that above I have written $\vec{A}'$ to denote the vector in the rotated coordinate system. Strictly speaking, this is an abuse of notation - since I am actually referring to the *same* vector, only in a different coordinate system[7]. I am sure that no harm will be done by this little mis-notation, though.

Continuing with this notation, let us now ask whether something that we have always called a scalar is really a scalar. I am referring to the scalar product of two vectors, $\vec{A}\cdot\vec{B}$. The question is, is $\vec{A}'\cdot\vec{B}' = \vec{A}\cdot\vec{B}$? To see that it is, we can play a little trick. Since we have already shown that lengths of all vectors do not change under a rotation, we have $\left(\vec{A}' + \vec{B}'\right)^2 = \left(\vec{A} + \vec{B}\right)^2$, which means that

$$\vec{A}'^2 + 2\,\vec{A}'\cdot\vec{B}' + \vec{B}'^2 = \vec{A}^2 + 2\,\vec{A}\cdot\vec{B} + \vec{B}^2$$

- using $\vec{A}'^2 = \vec{A}^2$ and $\vec{B}'^2 = \vec{B}^2$ immediately gives

$$\vec{A}'\cdot\vec{B}' = \vec{A}\cdot\vec{B}$$

So, the scalar product is really a scalar! Of course, you could have checked this directly from the transformation law itself.

Let me point out that the fact that the dot product of two vectors $\vec{A}\cdot\vec{B} = A_xB_x + A_yB_y + A_zB_z$ is a scalar depends on the fact that the transformation that I am considering preserves vector lengths. Such transformations are called *orthogonal* transformations. Rotations, of course, are orthogonal transformations in three dimensions - other well known transformations in this category are reflections. Indeed, it can be shown that in three dimensions, all

[7]There is a different way to think of a rotation. Instead of the vector staying the same and the coordinate system rotating, we can also keep the coordinate axes the same and turn the vector around the other way. In this case, the components change in the same fashion as above, and the notation $\vec{A}'$ is justified (we are really talking of a different vector here). Technically, the kind of transformations that we have been talking about are called *passive* transformations, while this latter kind are called *active* transformations.,

orthogonal transformations are either rotations, reflections or combinations of the two.

What if the basic transformation had been something else - one that does not preserve lengths? You may think it unnecessary to waste time on such grotesque possibilities - but the fact of the matter is that such transformations are not very uncommon. Indeed, the Lorentz transformations, as we will see very soon - fall in this bracket! In this case, we will have to modify the way in which we define the scalar product - so that it stays a scalar!

Now that we have a precise definition of vectors and scalars, you may realise that many quantities that we had assumed to be a vector may or may not be - the bottomline is - you have to check whether the transformation law is obeyed. For example, it is easy to check that the sum of two vectors is really a vector, *i.e.* it, too, transforms like the position vector (indeed, I should have checked whether this is correct *before* using this result in my proof of the invariance of the dot product above) and the same goes for the product of a vector with a scalar. As for the cross product of two vectors, I leave it to you to verify that this does transform like a vector, hence justifying its name.

5.5.2 Lorentz transformations vs. rotations

Just what is so special about rotations? Not much - if you are looking at things from a mathematician's point of view! A mathematician would be happy to think about all sorts of other linear transformations one can carry out on coordinates[8]. To him, a vector is a set of three components that change under any linear

[8]A linear transformation is one in which x' is a linear combination of x, y, z and so on. Thus, $x' = 3.6x - 1.7y + 0.5z$ is a linear transformation - while $x' = 2x^2$ is not.

transformation of coordinates[9] in the same way as the coordinates themselves. There is a special reason why rotations are special to a physicist, though! The laws of physics are the same no matter whether you write them with respect to one coordinate system or another, rotated system. In other words, rotations are symmetries of physical laws. This is precisely why we are so interested in what happens to objects under rotations.

Rotations are symmetries of all physical laws - but that can be traced back to the fact that they are special cases of a much more general symmetry - the Lorentz transformations. After all, the basic tenet of relativity is that all physical laws are the same for all observers - which means that the Lorentz transformations are symmetries of all physical laws. You will agree that we must look for quantities that transform in a decent way under Lorentz transformation.

5.5.3 Vectors and scalars in spacetime

As I have stressed over and over again, the Lorentz transformations play the same role in spacetime as the rotations in space. This similarity helps us extend the concept of vectors in space to vectors in spacetime. In perfect analogy with the ordinary vectors, we can define vectors in four dimensional spacetime as *a set of four numbers* (A^0, A^1, A^2, A^3) *that transform under a Lorentz transformation in the same manner as the spacetime coordinates* (x^0, x^1, x^2, x^3). In other words all vectors in spacetime transform in the same manner and the spacetime coordinates give their prototype. To avoid

[9]Indeed, the mathematician (and sometimes, the physicist as well!) would be equally at home talking about general (nonlinear) transformation of coordinates. In this case, the new coordinate differentials dx', dy', dz' are linear combinations of dx, dy and dz. This allows vectors to be defined as a set of 3 components that transform in the same way as the *coordinate differentials* do!

confusion with our ordinary run of the mill vectors we give these entities a special name - **four-vectors**. If I talk of these vectors and our familiar old vectors together, as I will do often, I will refer to the latter as three-vectors. Remember, though, mathematically there is not much of a difference between four-vectors and three-vectors, they are just vectors under Lorentz transformations and rotations, respectively.

As for scalars, there generalization to spacetime is even more straightforward. The ordinary scalars that we dealt with in three dimensions are single numbers do not change under rotations. Four-scalars are single numbers that do not change under Lorentz transformations. Note that because rotations also form special cases of Lorentz transformations, all four-scalars are three-scalars, too. The converse is not true, though - time is a three-scalar but does not stay invariant under a general Lorentz transformation. The same goes for the length of the position vector.

The spacetime four vector (ct, x, y, z) naturally splits up into a time part - the zero-th component, and the three other components that form the spatial part. In the same way, every four-vector (A^0, A^1, A^2, A^3) can be split up into a timelike part - the A^0, and a spacelike part, $\vec{A} = (A^1, A^2, A^3)$. The real significance of this split up can be seen when you realize that a rotation in three dimensional space is a special case of Lorentz transformations in spacetime - one in which the (x, y, z) get mixed up together, but t stays invariant. Under a rotation, too, a four-vector transforms in the same way as $(ct, \vec{r})$ - so that A^0 does not change and $\vec{A}$ turns around! Thus, the time part A^0 of a four-vector is a scalar under rotations, while the space part $\vec{A}$ is a three-vector under rotations!

Just as in the case of spacetime coordinates, the four-vector that I have been describing here is written with the index as a superscript. In the same vein, we call such vectors contravariant four-vectors.

In case the foregoing seems too abstract, let's take a look at the way four-scalars and four-vectors behave in two very simple case - the special Lorentz boost, and a rotation about the Z axis. To refresh your memory, let me repeat the transformation equations for the special Lorentz transformation in terms of spacetime coordinates $(ct, \vec{r})$, equations (5.1-a-5.1-d)

$$\begin{aligned} x'^0 &= \gamma x^0 - \beta\gamma x^1 \\ x'^1 &= -\beta\gamma x^0 + \gamma x^1 \\ x'^2 &= x^2 \\ x'^3 &= x^3 \end{aligned}$$

Under this transformation, a four scalar changes to

$$\phi \to \phi' = \phi$$

- *i.e.* it does not change at all. What about a four-vector with components $\left(A^0, \vec{A}\right)$? It transforms into

$$A'^0 = \gamma A^0 - \beta\gamma A^1 \tag{5.22-a}$$
$$A'^1 = -\beta\gamma A^0 + \gamma A^1 \tag{5.22-b}$$
$$A'^2 = A^2 \tag{5.22-c}$$
$$A'^3 = A^3 \tag{5.22-d}$$

What about the rotation? The spacetime transformation in this case is

$$x'^0 = x^0$$

$$\begin{aligned} x'^1 &= x^1 \cos\vartheta + x^2 \sin\vartheta \\ x'^2 &= -x^1 \sin\vartheta + x^2 \cos\vartheta \\ x'^3 &= x^3 \end{aligned}$$

Under this rotation, a four-scalar will obviously stay the same. What about the four-vector?

$$\begin{aligned} A'^0 &= A^0 \\ A'^1 &= A^1 \cos\vartheta + A^2 \sin\vartheta \\ A'^2 &= -A^1 \sin\vartheta + A^2 \cos\vartheta \\ A'^3 &= A^3 \end{aligned}$$

As promised, the component A^0 stays invariant under the rotation - whereas the other three components change in just the same way as the components of a three vector does. Of course, for this particular rotation, the component A_3 also stays unchanged, but that is because the rotation I have considered here is one about the third (Z) axis.

It is now time to go back to the general case once again. I will adopt a convention that is used almost universally - which is using Greek letters μ, ν etc., which take the values $0, 1, 2, 3$ to stand for the components of four-vectors. Very often, I will write A^μ to indicate not the μ-th component of the four-vector A but the four-vector itself - it should be clear from the context which one is being meant at a given point. Another convention that is used almost universally is to use Latin indices i, j, k etc., which run over the values of 1,2 and 3 to stand for the spatial component of a four vector.

5.5.4 Examples of 4-vectors and 4-scalars

Now that we know how to generalize the concept of vectors and scalars to spacetime, the natural question to ask is - are these just mathematical entities or do we have real life examples of these objects? Of course we do! That is the whole point behind studying them in the first place.

We have already met the simplest non-trivial scalar quantity - the interval! The by now familiar quantity

$$\left(x^0\right)^2 - \left(x^1\right)^2 - \left(x^2\right)^2 - \left(x^3\right)^2 \equiv c^2t^2 - \vec{r}^2$$

is invariant under Lorentz transformations. This is exactly what we mean by a four-scalar! The proper time interval between two timelike related events is defined by

$$\Delta\tau = \sqrt{\Delta t^2 - \frac{1}{c^2}\left(\Delta\vec{r}\right)^2} = \Delta t\sqrt{1 - \frac{1}{c^2}\left(\frac{\Delta\vec{r}}{\Delta t}\right)^2}$$

is also obviously a four scalar.

The most obvious four-vector is none other than the coordinate four vector $x^\mu = (ct, \vec{r})$. We have already seen that differences in spacetime coordinates change in the same way when you jump from one inertial frame to another as the coordinates themselves - so another four-vector is$\Delta x^\mu = (c\Delta t, \Delta\vec{r})$.

What about other four-vectors? Since the three-vector part of a four vector has to be a vector under rotations, it might seem that all you have to do to build more four-vectors of your own is to take an ordinary vector and add to it an appropriate three-scalar to serve as the zero-th (or temporal) component. That this does not work can be seen immediately. Take the simplest three-vector beyond the position vector that you can think of in mechanics - the velocity vector $\vec{u}$. Try as you might, you can't find a three scalar ζ that you can club with this one to get a four-vector. The easiest way to see this is to consider the special Lorentz boost. If $(\zeta, \vec{u})$ is a four vector, then under this the velocity components in the Y and Z direction must stay unchanged - which is not what our velocity transformation equations (3.17-b) and (3.17-c) say!

To understand how to proceed further, then, we have to take a deeper look at what we had always thought obvious - just why is velocity a three-vector? Well $\vec{r}$ is a three-vector and hence so is $\Delta\vec{r}$. The time interval Δt is a three scalar, so that the ratio $\Delta\vec{r}$ is really a product of a scalar $\frac{1}{\Delta t}$ and a vector $\Delta\vec{r}$ - and such a product does transform like the coordinates under a rotation. Velocity is nothing but the limit of the ratio $\frac{\Delta\vec{r}}{\Delta t}$ as $\Delta t \to 0$, so it is a limit of a sequence of vectors. This is why velocity is a three-vector. As you can see, the crucial point above is that Δt is a scalar under rotations.

You can try the same trick with the position four vector x^μ and land up with $\frac{\Delta x^\mu}{\Delta t} = (c, \vec{u})$. Why is this not a four-vector? The answer is simple - while Δt was a three-scalar under rotations, it is not a scalar under the Lorentz transformations. So, if you want to get the four dimensional version of the velocity three-vector, you have to look for something other than Δt - something that is similar to it (otherwise we will land up with an entirely different object - not a generalization of the velocity!) and is invariant under a Lorentz transformation. Fortunately, such a thing is not very difficult to find - the proper time $\Delta\tau$ is the perfect candidate! Not only is it an invariant, it has the same dimensions as time and reduces (approximately) to it when the speed $\left|\frac{\Delta\vec{r}}{\Delta t}\right|$ is small. This helps us to define the **velocity four-vector** by

$$U^\mu \equiv \frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt} \times \frac{dt}{d\tau} = \left(\frac{c}{\sqrt{1-\frac{u^2}{c^2}}}, \frac{\vec{u}}{\sqrt{1-\frac{u^2}{c^2}}} \right) = \gamma\left(c, \vec{u}\right) \tag{5.23}$$

where I have used

$$\frac{d\tau}{dt} = \lim_{\Delta t \to 0} \frac{\Delta\tau}{\Delta t} = \lim_{\Delta t \to 0} \sqrt{1 - \frac{1}{c^2}\left(\frac{\Delta\vec{r}}{\Delta t}\right)^2} = \sqrt{1-\frac{u^2}{c^2}}.$$

Note that the three-vector part $\vec{U} = \gamma\vec{u}$ of our velocity four-vector is precisely the vector that I had introduced back in subsection 3.5.1.4. It should now be obvious why the components U_y and U_z were transforming that simply under the special Lorentz transformation! Indeed, the identity (3.27) that requires a somewhat messy bit of algebra to prove directly, becomes nearly trivial once you realise that the quantity $c\gamma = \frac{c}{\sqrt{1-\frac{u^2}{c^2}}}$ forms the zero-th component of a four vector. So, it would transform in the same way as ct does under the special Lorentz transformation, leading to

$$\frac{c}{\sqrt{1-\frac{u'^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left(\frac{c}{\sqrt{1-\frac{u^2}{c^2}}} - \frac{v}{c}\,\frac{u_x}{\sqrt{1-\frac{u^2}{c^2}}} \right) = \frac{c\left(1-\frac{u_x v}{c^2}\right)}{\sqrt{1-\frac{v^2}{c^2}}\sqrt{1-\frac{u^2}{c^2}}}$$

from which (3.27) follows immediately! By the way, this is a special version of the identity - valid when the relative velocity of the two frames is along the X axis. Generalising to an arbitrary direction of the relative velocity becomes obvious when you realise that most of the terms that are there in the identity are manifestly three-scalars (remember - this means that they do not change under a rotation of the axis), like v^2, u^2 etc. The only one that is not manifestly a three-scalar is the term $u_x v$. However, once you realise that here the vector $\vec{v}$ is $(v,0,0)$, it is easy to see that $u_x v = \vec{u}\cdot\vec{v}$, and hence this identity becomes, in general

$$\frac{\gamma(v)\,\gamma(u)}{\gamma(u')} = \frac{\sqrt{1-\frac{v^2}{c^2}}\sqrt{1-\frac{u^2}{c^2}}}{\sqrt{1-\frac{u'^2}{c^2}}} = 1 - \frac{\vec{u}\cdot\vec{v}}{c^2}. \tag{5.24}$$

To round off our small discussion on the velocity four-vector, let me do what I have done a lot of times before - show you yet another way of deriving the same old results. This time, I am going to use that fact that U^μ is a four vector to derive the transformation laws for the ordinary velocity. It is easy to see the transformation laws for the spatial part of the four-vector U^μ are

$$\begin{aligned}
U'^1 \equiv \frac{u'_x}{\sqrt{1-\frac{u'^2}{c^2}}} &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left(U^1 - \frac{v}{c}U^0\right) = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left(\frac{u_x}{\sqrt{1-\frac{u^2}{c^2}}} - \frac{v}{c}\frac{c}{\sqrt{1-\frac{u^2}{c^2}}}\right)\\
U'^2 \equiv \frac{u'_y}{\sqrt{1-\frac{u'^2}{c^2}}} &= U^2 = \frac{u_y}{\sqrt{1-\frac{u^2}{c^2}}}\\
U'^3 \equiv \frac{u'_z}{\sqrt{1-\frac{u'^2}{c^2}}} &= U^3 = \frac{u_z}{\sqrt{1-\frac{u^2}{c^2}}}
\end{aligned}$$

dividing these equations by the transformation equation for the zero-th component above immediately leads to the familiar velocity transformation equations.

As you must have guessed, there is no reason to stop at the velocity four-vector. You can continue in the same vein and define the acceleration four-vector $A^\mu \equiv \frac{dU^\mu}{d\tau}$. This time you have to be a bit careful, though - the factor γ that relates the velocity four-vector to the ordinary velocity is a variable, too, leading to the connection between the acceleration four-vector and its three-vector counterpart a bit more complicated

$$\begin{aligned} A^\mu &= \frac{dU^\mu}{d\tau} = \gamma \frac{dU^\mu}{dt} \\ &= \gamma \frac{d}{dt}\left[\gamma\left(c, \vec{u}\right)\right] \\ &= \gamma \left[\frac{d\gamma}{dt}\left(c, \vec{u}\right) + \gamma\left(0, \vec{a}\right)\right] \end{aligned}$$

Now $\frac{1}{\gamma}\frac{d\gamma}{dt} = \frac{d}{dt}\left(\ln \gamma\right) = \frac{u}{c^2}\gamma^2\frac{du}{dt} = \frac{\gamma^2}{c^2}\vec{u}\cdot\vec{a}$, so that

$$A^\mu = \gamma^2\left[\frac{\gamma^2}{c^2}\vec{u}\cdot\vec{a}\left(c, \vec{u}\right) + \left(0, \vec{a}\right)\right] = \gamma^2\left(\frac{\gamma^2}{c}\vec{u}\cdot\vec{a}, \vec{a} + \frac{\gamma^2}{c^2}\left(\vec{u}\cdot\vec{a}\right)\vec{u}\right) \tag{5.25}$$

which shows that even the space part of the acceleration four-vector is not necessarily even in the same direction as the acceleration $\vec{a}$! This result is going to have quiet an important bearing on relativistic mechanics - the subject of the next chapter.

5.5.5 The dot product of four vectors

A very standard way of getting a scalar from two vectors in three dimensions is taking their dot product. We have already verified above that this really is a scalar! Can we do such a thing in four dimensions? Yes, we can - but there is a slight twist. If I just take the formula for the three dimensional dot product

$$\vec{A}\cdot\vec{B} = A_1B_1 + A_2B_2 + A_3B_3$$

and try writing down its four dimensional version what I will land up with is not a scalar. This is easy to see - just try it out on our prototype four vector $(ct, \vec{r})$ and you will get $c^2t^2 + \vec{r}^2$, which is not invariant under a Lorentz transformation! Fixing this up is not a very big problem, though - one look at the invariant interval suggests that instead of the sum we should try

$$A \cdot B = A^0 B^0 - A^1 B^1 - A^2 B^2 - A^3 B^3$$

In terms of the metric tensor η that I had introduced you to earlier, this can be easily written in the form

$$A \cdot B = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} A^\mu B^\nu \tag{5.26}$$

or, in terms of column vectors $A = \left(A^0, A^1, A^2, A^3\right)^T$ and $B = \left(B^0, B^1, B^2, B^3\right)^T$ and the matrix η

$$A \cdot B = A^T \eta B \tag{5.27}$$

It is rather simple to check that this product is actually an invariant. In the matrix representation, since $A \to A' = LA$ and $B \to B' = LB$, we get

$$\begin{aligned} A \cdot B \to A' \cdot B' &= A'^T \eta B' \\ &= (LA)^T \eta LB \\ &= A^T L^T \eta LB \\ &= A^T \eta B = A \cdot B \end{aligned}$$

where I have used equation (5.13). Another important property of the scalar product of four-vectors, that it shares with the three-vector dot product is that it is symmetric

$$A \cdot B = B \cdot A \tag{5.28}$$

In direct analogy with the covariant spacetime components x_μ that I had introduced earlier, we can define a covariant four-vector

$\underline{\mathrm{A}}$ corresponding to a contravariant four vector A by

$$\underline{\mathrm{A}} = \eta A \tag{5.29}$$

or, in terms of components

$$A_\mu = \sum_{\nu=0}^{3} \eta_{\mu\nu} A^\mu \tag{5.30}$$

Note that though I have used the underbar sign in the matrix form of the covariant vector to distinguish it from the covariant version, this is not necessary for the components - the placement of the index μ (as a subscript rather than a superscript) is enough to make the distinction.

In terms of the covariant version of a vector it is easy to see that one can write a compact version of the formula for the dot product

$$A \cdot B = \sum_{\mu=0}^{3} A^\mu B_\mu = \sum_{\mu=0}^{3} A_\mu B^\mu \tag{5.31}$$

This formula should be strongly reminiscent of the formula for the 3 vector dot product

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^{3} A_i B_i$$

Note that in this case you had to take the corresponding components of the two vectors, multiply them and add up all the terms together to get a scalar. According to (5.31), you do exactly the same for the four-vector dot product, except that in this case the two vectors have to be of different kinds - one contravariant and the other covariant! Another way to put this is - you can't produce a scalar by using the multiply coefficients and add rule on two contravariant vectors (or two covariant vectors either) - for this rule to work, you must involve the two different kinds of vectors!

To understand exactly why two kinds of vectors are needed, let us try to prove directly the invariance of the product in (5.31) in terms of the Lorentz transformation matrices. It is easy to see that in terms of the column vectors $\underline{A} = (A_0, A_1, A_2, A_3)^T$ and $B = (B^0, B^1, B^2, B^3)^T$ the inner product can be written as

$$A \cdot B = \underline{A}^T B.$$

Now, under a Lorentz transformation, we have $B \to B' = LB$ but $\underline{A} \to \underline{A}' = \underline{L}\,\underline{A}$. This means that

$$\underline{A}^T B \to \underline{A}'^T B' = (\underline{L}\,\underline{A})^T LB = \underline{A}^T \underline{L}^T LB = \underline{A}^T B$$

where the last step follows from the fact that $\underline{L} = \left(L^T\right)^{-1}$. Note that had you tried to do the same same with two contravariant vectors, the matrix product $\underline{L}^T L$ would have been replaced by $L^T L$ - which would have not been equal to the identity matrix.

This should also tell you why you had not met two kinds of vectors before, when you were studying them in three dimensions. There, the ransformations you were talking of were rotations, and there is no distinction between the transformation matrix R and $\underline{R} = \left(R^T\right)^{-1}$. So, it all boils down to the fact that rotations are orthogonal (which makes $R = \left(R^T\right)^{-1}$), while Lorentz transformations are not.

> One reason why the Minkowski coordinates that I introduced in subsection 5.4.3 were useful is that they made the Lorentz transformations orthogonal! This meant that in terms of these coordinates, you did not have to distinguish between contravariant and covariant vectors. This should explain why we used subscripts on the coordi-

nates x_1, x_2, x_3, x_4 - since there is no need to distinguish between two kinds of entities, we stick to the more conventionl choice! As I discussed in subsection 5.4.3, the conceptul difficulties involved with an imaginary time coordinate tends to outweigh these small advantages.

One of the most common example of a dot product is that of a vector with itself, written A^2 - which yields, as in the three dimensional case, the four dimensional invariant norm of the vector. In the three dimesional case, the square root of this quantity is the length of the vector. This does not make much sense, though, for the four dimensional case - since the mixed signature of the metric allows the norm of a vector to be negative as well as positive or zero. Note that the norm of a nonzero vector *can* be zero!

The norm of our prototype four-vector, $(ct, \vec{r})$ is the invariant interval between the event at $(ct, \vec{r})$ and the spacetime origin $\left(0, \vec{0}\right)$

$$(ct, \vec{r})^2 = c^2 t^2 - \vec{r}^2$$

It is easy to see that the intreval between two coordinates with spacetime four-vectors x_1 and x_2 is given by $(x_2 - x_1)^2$. In analogy to the way in which we classify coordinate four-vectors as timelike, spacelike and lightlike in terms of the sign of its norm, we calssify all four-vectors as either temporal, spatial or null according as its norm is positive, negative or zero.

Lets take a look at the other four-vectors we have met in the last section. The norm of the velocity four-vector $\gamma (c, \vec{u})$ is given by

$$U^2 = \gamma^2 c^2 - \gamma^2 u^2 = \gamma^2 \left(c^2 - u^2\right) = c^2 \tag{5.32}$$

Thus the norm of the four-velocity is an invariant - though, being constant, it does not seem to be a very interesting one at this stage!

As can be see, the velocity four vector of any particle has to be timelike.

What about the acceleration four-vector,

$$A^\mu = \gamma^2 \left(\frac{\gamma^2}{c} \vec{u} \cdot \vec{a}, \vec{a} + \frac{\gamma^2}{c^2} (\vec{u} \cdot \vec{a}) \vec{u} \right) \quad ?$$

A bit of messy algebra shows that its norm is given by

$$A^2 = -\gamma^4 \left(\vec{a}^2 + \frac{\gamma^2}{c^2} (\vec{u} \cdot \vec{a})^2 \right) \tag{5.33}$$

You can, if you are a glutton for punishment, verify directly that this is invariant under a Lorentz transformation (or at least, under the special Lorentz transformation)! One thing that follows directly from the above is that the acceleration four-vector is always space-like, unless of course the three-acceleration $\vec{a}$ is zero, in which case it is lightlike[10].

Let me show you another way of deriving the result above. At any given instant the particle has a given velocity $\vec{u}$ with respect to the inertal observer being used. Let us now move over to another *inertial* observer, to whom the particle is at rest at that instant. Note that if the particle has a non-zero acceleration, it is going to start moving in that frame just afterwards. You can always find a frame in which the particle stays at rest always - all you have to do is put an observer right on the particle. Such an observer will be non inertial, and we won't have much to do with these creatures in STR. What we have to be satisfied is the *instantaneously comoving frame* (ICF) of the kind described above. In that frame the acceleration of the particle is often called its proper accelera-

[10] That's hardly surprising, of course, given that for $\vec{a} = 0$ the acceleration four-vector is identically zero!

tion $\vec{a}_0$, and since $\vec{u} = 0, \gamma = 1$ the acceleration four-vector simplifies significantly to $(0, \vec{a}_0)$. Thus in a particle's ICF, the norm of its four-acceleration is given by $-\vec{a}_0^2$ - which is obviously negative (or zero, if $\vec{a}_0 = \vec{0}$)! Now comes the crux of the argument - since the norm of a four-vector is invariant, any other observer has to measure it as $-\vec{a}_0^2$ - so that everyone agrees that it is negative or zero!

The trick that I have used here is one that is used quiet often in relativistic calculation. The fact that scalars are the same for all inertial observers allow us to jump from one inertial observer to another, for whom the calculation can be simple. In this context, the rest frame of a particle (which has to be replaced by its ICF(s) if it is accelerating) is very often the one in which the calculation becomes simplest. The theme is - calculate once, use for all! The same can be applied, to a somewhat lesser extent, to four-vector quantities, too (and as we will see later, to other kinds of stuff, like four-tensors and so on). The only trouble is, in this case you have to take the trouble of transforming back to your original frame of referenc once the calculation is done. Actually, you had to do this for scalars too - the only difference being that this is trivial for scalars.

To illustrate the power of technique let me use it to derive an expression that will be of some importance in the study of relativistic mechanics - the value of the scalar $A \cdot U = A^\mu U_\mu$. In the ICF of a particle, these two vectors take values $A = (0, \vec{a}_0)$ and $U = \left(c, \vec{0}\right)$ respectively. It is easy to see that in the ICF,

$$A \cdot U = 0 \times c - \vec{a}_0 \cdot \vec{0} = 0 \tag{5.34}$$

and this means that this quantity vanishes for all frames. You can (and should) check that this is true by doing the explicit calculation using (5.23) and (5.25) - which should convince you of the utility of this trick.

There is another way in which the result (5.34) could be derived - and the method is quiet instructive in its own right. It is easy to check that the familiar calculus law for the derivative of a product

$$\frac{d}{dt}\left(uv\right) = u\frac{dv}{dt} + \frac{du}{dt}v$$

can be extended to the scalar product almost in toto

$$\frac{d}{d\tau}\left(A\cdot B\right) = \frac{dA}{d\tau}\cdot B + A\cdot\frac{dB}{d\tau} \tag{5.35}$$

To use this in our present task, let's start with the norm of the four-velocity

$$U^2 = U\cdot U = c^2$$

and differentiate to get

$$U\cdot\frac{dU}{d\tau} + \frac{dU}{d\tau}\cdot U = 0$$

Using the definition $A^\mu \equiv \frac{dU^\mu}{d\tau}$ and the symmetry of the dot product, this translates to $U\cdot A + A\cdot U = 2\,A\cdot U = 0$, leading to our identity[11].

As a final example of the utility of this trick of evaluating four-scalar quantities in a special frame, secure in the knowledge that their value will not be affected when you move back to a general frame of reference, I will show you yet another way of deriving (**??**)- this time almost trivially. Note that for the four-velocities

[11]You may recall a similar result from the world of three-vectors - the derivative of any vector of fixed length is perpendicular to it - that is proved in exactly the same way! Among its consequence is the fact that the tangent of a circle is perpendicular to the radius, as well as the fact that the acceleration of a particle in uniform circular motion is centripetal (since it is perpendicular to the velocity).

$V = \gamma(v)(c, \vec{v})$ and $U = \gamma(u)(c, \vec{u})$ the four-scalar product is given by

$$U \cdot V = \gamma(u)\gamma(v)\left(c^2 - \vec{u} \cdot \vec{v}\right)$$

while its value in the rest frame of the second observer, where the particle moves with the relative velocity $\vec{u}'$ is,

$$U \cdot V = \gamma(u')(c, \vec{u}') \cdot \left(c, \vec{0}\right) = c^2\gamma(u').$$

Equating these two results for $U \cdot V$ gives us (5.24) immediately!

5.5.6 Going further - tensors

Now that we have understood the most basic physical entities that you can find in spacetime, it is time to go a bit further and ask - what other kinds of physical entities can there be? The most obvious extenion to the concepts of four-scalars and four-vectors are four-tensors. Indeed, as you may know, scalars and vectors can even be thought of as special cases of tensors.

To understand tensors in four dimensions, let us first take a look at a familiar tensor from the three dimensional world. I am almost certain that the most familiar tensor that you can think of is - the moment of inertia tensor I. It is what relates the angular velocity $\omega$of a body to its angular momentum L,

$$L = I\omega \tag{5.36}$$

where the moment of inertia tensor is represented by a 3×3 matrix, with the vectors L and ω being represented by 3×1 column vectors. From the way we have been defining scalars and vectors, it should be obvious that the question to ask is - just how does the moment of inertia tensor transform? We can easily deduce this from the vector transformation law. Simply note that

$$L \to L' = RL, \qquad \omega \to \omega' = R\omega$$

which implies that the moemnt of inertai tensor in the rotated coordinate frame must satisfy

$$RL = I'R\omega, \quad \Rightarrow \quad RI\omega = I'R\omega \qquad \text{for all } \omega$$

which means that we must have

$$I' = RIR^{-1} = RIR^T \tag{5.37}$$

where in the last line I have used the fact that the rotation matrix is orthogonal. This is the law of transformation for the moment of inertia tensor[12]. A bit of thought should show you that the same law should work for anything that can be written in the form of a square matrix that describes the linear relationship between one vector and another. Thus, the dielectric tensor (relating the electric field vector $\vec{E}$ to the electric displacement $\vec{D}$), the stress tensor (relating the area vector to the force vector), etc. all transform according to this rule. In general, the class of objects that transform like this are collectively called tensors of rank two.

Why "rank two"? Because the rotation matrix appears twice (once as R and once as R^T) in the transformation law (5.37)! This may be even more obvious if I write this transformation law in

[12]In the language of matrices, this is an orthogonal transformation, a special kind of similarity transformation in which the transforming matrix happens to be orthogonal. A theorem of linear algebra, which states that any symmetric matrix can always be orhogonally diagonalised, ensures that you will always be able to rotate your coordinate system into one in which the moment of inertia tensor is diagonal. This, of course, is the principle axis system - something which makes the study of rigid body dynamics very convenient.

terms of the components. Then

$$\begin{aligned} I'_{ij} &= \left(RIR^T\right)_{ij} \\ &= \sum_{a,b=1}^{3} R_{ia} I_{ab} \left(R^T\right)_{bj} \\ &= \sum_{a,b=1}^{3} R_{ia} R_{jb} I_{ab} \end{aligned}$$

where the two R matrix elements becomes explicit - especially if you compare with the corresponding law for vectors

$$v'_i = \sum_{a=1}^{3} R_{ia} v_a$$

A second rank tensor, then, can be *defined* as a set of $3 \times 3 = 9$ components that transform under a coordinate rotation according to the rule

$$T'_{ij} = \sum_{a,b=1}^{3} R_{ia} R_{jb} T_{ab}, \qquad i, j = 1, \ldots, 3 \tag{5.38}$$

This form makes it easy to see how to generalize this to get the third rank tensor - this is nothing but the set of 3^3 components that transform according to the rule

$$T'_{ijk} = \sum_{a,b,c=1}^{3} R_{ia} R_{jb} R_{kc} T_{abc}, \qquad i, j, k = 1, \ldots, 3 \tag{5.39}$$

You can keep on going higher and higher in rank by adding more indices and more R matrix coefficients in the transformation law!

All these was about tensors in three dimensions - or more precisely, tensors under rotations. What is the scenario in four dimensional spacetime? Our tensor of rank two (undrer rotations) was something that acts linearly on a vector to yield another vector. You should immediately see that something more complicated is going on here - the fact that there is more than one kind of vector in spacetime immediately tells us that there should be an even larger number of kinds of tensors of rank two, which change

- A covariant vector to a contravariant vector
- A contravariant vector to a covariant vector
- A covariant vector to a covariant vector
- A contravariant vector to a contravariant vector

, respectively!

Let me focus on a tensor that converts a covariant vector to a contravariant one for the time being - the other cases are similar. So, we have

$$V = T\,\underline{\mathrm{W}}$$

Since under a Lorentz transformation $x \to x' = Lx$, the two vectors transform according to $V \to V' = LV$ and $\underline{\mathrm{W}} \to \underline{\mathrm{W}}' = \left(L^T\right)^{-1}\underline{\mathrm{W}}$, we see that T must transform to T' given by

$$LT\underline{\mathrm{W}} = T'\left(L^T\right)^{-1}\underline{\mathrm{W}}$$

for all $\underline{\mathrm{W}}$, which readily shows that

$$T \to T' = LTL^T$$

which becomes, in terms of components

$$T'^{\mu\nu} = \sum_{\rho,\sigma=0}^{3} L^\mu{}_\rho L^\nu{}_\sigma T^{\rho\sigma} \tag{5.40}$$

As you can see, this tensor has two L matrix elements in its transformation, that's twice as many as a contravariant vector. It is easy to see, then, why it is called a contravariant tensor of rank . This also explains why I chose to put the indices up as superscripts in this case.

You should be able to show quiet easily that a tensor that changes a contarvariant vector into a covariant one transforms according to the law

$$T'_{\mu\nu} = \sum_{\rho,\sigma=0}^{3} L_{\mu}^{\ \rho} L_{\nu}^{\ \sigma} T_{\rho\sigma} \tag{5.41}$$

and is called a covariant tensor of rank 2. As for the other two possibilities, I will leave you to figure out that they yield mixed tensors with one contravariant and one covariant index, transforming as

$$T'^{\mu}_{\ \nu} = \sum_{\rho,\sigma=0}^{3} L^{\mu}_{\ \rho} L_{\nu}^{\ \sigma} T^{\rho}_{\ \sigma}. \tag{5.42}$$

One way in which you can directly produce a tensor of rank wo from two vectors is by forming the so called direct product. As an example, I will start with two contravariant 4-vectors, A^{μ} and B^{μ}, and form the set of 16 quantities

$$T^{\mu\nu} = A^{\mu} B^{\nu}, \qquad \mu, \nu = 0, 1, 2, 3 \tag{5.43}$$

It is very easy to check that this does satisfy the defining transformation property of a contravariant tensor of rank two :

$$\begin{aligned} T'^{\mu\nu} &= A'^{\mu} B'^{\nu} \\ &= \left(\sum_{\rho=0}^{3} L^{\mu}_{\ \rho} A^{\rho} \right) \left(\sum_{\sigma=0}^{3} L^{\nu}_{\ \sigma} B^{\sigma} \right) \end{aligned}$$

$$= \sum_{\rho,\sigma=0}^{3} L^{\mu}{}_{\rho} L^{\nu}{}_{\sigma} A^{\rho} B^{\sigma}$$
$$= \sum_{\rho,\sigma=0}^{3} L^{\mu}{}_{\rho} L^{\nu}{}_{\sigma} T^{\rho\sigma}$$

A word of warning - although the direct product of two vectors does give rise to a tensor of rank two - all tensor of rank two can not be written as a direct product of two vectors. A simple counting arguments should be enough to underatnd this - two vectors give you 8 numbers to play with - while a general tensor of rank two has 16 components! It is easy to see that the sum of the direct products $A^{\mu}B^{\nu}$ and $C^{\mu}D^{\nu}$ is a tensor of rank two, but it can be itself written as a direct product only if one of the vectors C, D is proportional to one of A, B.

A particularly important second rank tensor that you can form from two vectors U and V is the antisymmetrized direct product

$$A^{\mu\nu} = U^{\mu}V^{\nu} - U^{\nu}V^{\mu} \tag{5.44}$$

It is easy to see that if you restrict yourself to the spatial components $A^{ij},\ i,j = 1,2,3$ you get three independent numbers A^{23}, A^{31} and A^{12} which are old friends, three components of the cross product $\vec{U} \times \vec{V}$. Thus, (5.44) above is 4 dimensional generalization of the cross product of two three-vectors. Apart from the three spatial components that we have talked about earlier, the three other independent components are the time-space components $A^{0i},\ i = 1,2,3$. They form, obviously, the three components of the three-vector[13]

$$\vec{A} = U^0\vec{V} - V^0\vec{U}$$

[13]This is obviously a three-vector under rotations, since $\vec{U}$ and $\vec{V}$ are three-vectors and U^0 and V^0 are three-scalars.

- thus our antisymmetric product can be thought of as an entity that packs in two three vectors, the $\vec{A}$ above and $\vec{B} = \vec{U} \times \vec{V}$. Note that under rotations these two pieces mix among themselves - intermixing only happens under a more general Lorentz transformation, like a boost.

Since we will meet entities like this quiet often later, it may be a good idea to figure out how this transforms under, say, the special Lorentz transformation. Using the matrix form, we get

$$\begin{pmatrix} 0 & A_1' & A_2' & A_3' \\ -A_1' & 0 & B_3' & -B_2' \\ -A_2' & -B_3' & 0 & B_1' \\ -A_3' & B_2' & -B_1' & 0 \end{pmatrix} =$$
$$\begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & A_1 & A_2 & A_3 \\ -A_1 & 0 & B_3 & -B_2 \\ -A_2 & -B_3 & 0 & B_1 \\ -A_3 & B_2 & -B_1 & 0 \end{pmatrix} \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^T$$

The matrix multiplication above looks much more complicated than it is - and with a bit of perseverence you can easily show that this implies

$$A_1' = A_1 \tag{5.45-a}$$
$$A_2' = \gamma\left(A_2 - \beta B_3\right) \tag{5.45-b}$$
$$A_3' = \gamma\left(A_3 - \beta B_2\right) \tag{5.45-c}$$
$$B_1' = B_1 \tag{5.45-d}$$
$$B_2' = \gamma\left(B_2 + \beta A_3\right) \tag{5.45-e}$$
$$B_3' = \gamma\left(B_3 - \beta A_2\right) \tag{5.45-f}$$

This result will be quiet useful for us in the future.

It is easy to see that you can form tensors of other kinds by taking a direct product of vectors of the appropriate kind. Indeed, by playing this game with more than two vectors will give you a tensors of higher rank. For instance, taking the direct product of three contravariant vectors produces a tensor of rank three, a quantity that transfoms according to

$$T'^{\lambda\mu\nu} = \sum_{\vartheta,\rho,\sigma=0}^{3} L^{\lambda}_{\ \vartheta} L^{\mu}_{\ \rho} L^{\nu}_{\ \sigma} T^{\vartheta\rho\sigma} \tag{5.46}$$

Remember that, once again, a general third rank tensor will transform according to this equation but will not be, in general, a direct product of three vectors.

In general, it is easy to see that we can take the direct product of m contravariant vectors and n covariant vectors to produce an object that trasforms like

$$T'^{\mu_1\ldots\mu_m}{}_{\nu_1\ldots\nu_n} = \sum_{\substack{\rho_1,\ldots,\rho_m \\ \sigma_1,\ldots,\sigma_n}=0}^{3} L^{\mu_1}_{\rho_1}\ldots L^{\mu_m}_{\rho_m} L^{\sigma_1}_{\nu_1}\ldots L^{\sigma_n}_{\nu_n} T^{\rho_1\ldots\rho_m}{}_{\sigma_1\ldots\sigma_n} \tag{5.47}$$

Any set of 4^{m+n} numbers that transform like the above are called mixed tensors having a contarvariant rank of m and a contravariant tesor of rank n or, in short, a tensor with a rak of $m+n$.

5.5.7 Why bother?

We will meet four-scalars, four-vectors and four-tensors in a big way in the next few chapters. Are you wondering why they are so important? The reason lies in the principle of relativity. This implies that in order to qualify as a valid physical law a law has to be equally valid for all inertial frames. In other words, the equations that express these laws must be Lorentz covariant. Note the choice of words here - I am saying covariant, not invariant. What this means is that the two sies of an equation may change (not remain invariant), but change in the same manner (co-vary) under a Lorentz transformation - so that if the equation is correct for one inertial frame of reference, it is correct for any other inertial frame.

So, given any proposed law of physics, it becomes important to check whether it conforms to this principle of Lorentz covariance. Now, this may be a difficult matter - you will have to take a look at the way in which each entry on both sides of the equation changes under a Lorentz transformation, carry out the necessary algebra and often rather painfully verify that both sides do change in the same manner.

This is exactly where four-tensors come in. If you write down an equation in which both sides are four-tensors (of the same kind) then Lorentz covariance is guaranteed - both sides of the equation will change on changing the inertial observer, but will change in exactly the same way. The point is - if we write equations in terms of matching four dimensional entities, then there is no longer any need to check for Lorentz covaraince - such equations are *manifestly* Lorentz covariant!

Chapter 6

Kinetics in relativity

6.1 Rewriting Newton

6.2 The momentum - kinetic energy connection

Let me remind you how we defined kinetic energy in Newtonian mechanics. It is the work that has to be done on a stationary particle to speed it up from rest. Thus,

$$\begin{aligned} T &= \int \vec{F} \cdot d\vec{r} \\ &= \int \frac{d\vec{p}}{dt} \cdot d\vec{r} = \int_{\vec{0}}^{\vec{p}} \vec{u} \cdot d\vec{p} \end{aligned} \tag{6.1}$$

This formula remains valid in relativity, too. Note that in classical mechanics we have $\vec{p} = m\vec{u}$, which leads to the well known result

$$T = \frac{p^2}{2m} = \frac{1}{2}mu^2 \tag{6.2}$$

Note that for a photon, which travels always with the speed of light, the kinetic energy and the momentum are simply related by

$$T_\gamma = \int_0^{p_\gamma} c dp = p_\gamma c \tag{6.3}$$

We will borrow from quantum theory the result that the energy of a photon of frequency ν is given by $T_\gamma = h\nu$. Here h stands for Planck's constant. Then the photon must have a momentum given by $p_\gamma = \frac{h\nu}{c} = \frac{h}{\lambda}$. I will make extensive use of these results in the next few sections to show, firstly, that the classical definitions for momentum (and hence kinetic energy) are inadequate in relativity, and secondly, to find out what they should be redefined as!

I told you before that the Doppler factor is in many senses (except familiarity) a better measure of motion than the speed. Not only is it easier to determine, it also makes the algebra much simpler in relativistic calculations. In keeping with this, I will write both the momentum and kinetic energy of a particle as functions of its Doppler factor[1] k, namely, as $p(k)$ and $T(k)$. Classical physics tells us that

$$\begin{aligned} p(k) &\equiv mu = mc\frac{k^2-1}{k^2+1} \\ T(k) &\equiv \frac{1}{2}mu^2 = \frac{1}{2}mc^2\left(\frac{k^2-1}{k^2+1}\right)^2 \end{aligned}$$

Of course, the versions in terms of v looks a lot simpler, but that is only because these are non-relativistic formulae! As we will see in the next section, the demand that the law of momentum conser-

[1] More precisely, it is the Doppler factor of an observer that shares the particle's velocity in that instant of time. Note that the particle's k may change over time - which means that an observer fixed to it can not be inertial. Instead we may think of instantaneously comoving frames - a sequence of inertial frames in each of which the particles is at rest at that instant.

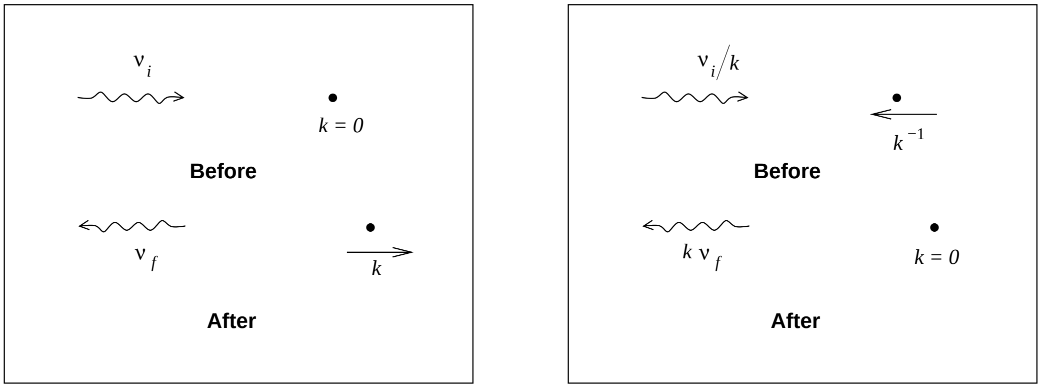


Figure 6.1: A photon bouncing off an electron

vation works in all inertial frames will tell us that these relations must be wrong! They are nevertheless very important - any correct formula *must* reduce to these when the velocity is small (*i.e.* k is close to unity) compared to the speed of light!

6.3 The need to redefine momentum

Let's try shining our flashlight at an electron at rest. I will consider a very simple situation. Here, the photon hits the electron and bounces straight back. In the process, it transfers some of its energy to the electron. Thus the photon comes out lower in energy and hence lower in frequency. The situation is shown in figure 6.1a.

Here, the initial net momentum is $\frac{h\nu_i}{c} + p\,(1)$ (remember that $k = 1$ for an electron at rest), while the final one is $-\frac{h\nu_f}{c} + p\,(k)$. As for kinetic energy, the initial and final values for it are $h\nu_i + T\,(1)$ and $h\nu_f + T\,(k)$, respectively. Since a particle at rest has neither momentum nor kinetic energy, we know that $p\,(1) = 0$ and $T\,(1) = 0$. This means that in the frame where the electron is initially at rest, the conservation of energy and momentum leads to the equations

$$\frac{h}{c}\left(\nu_i + \nu_f\right) = p\left(k\right) \tag{6.4}$$

$$h\left(\nu_i - \nu_f\right) = T\left(k\right) \tag{6.5}$$

In this problem, the two unknowns are ν_f, the final frequency of the photon, and k the final Doppler factor (*i.e.* speed) of the electron. We have two equations - so it seems we have all the ingredients needed to solve the problem. The only trouble is, we don't yet know the functions $p\left(k\right)$ and $T\left(k\right)$!

One result that follows immediately from (6.4) and (6.5) is that for the ratio between the kinetic energy and momentum

$$\frac{T\left(k\right)}{cp\left(k\right)} = \frac{\nu_i - \nu_f}{\nu_i + \nu_f} \tag{6.6}$$

- a formula that will be of some use later.

What does this collision look like from another frame - the one in which the electron is at rest *after* the collision? It is easy to see that in this case, the initial electron is moving to the left with a Doppler factor k^{-1} (same speed, opposite direction of motion!), so that the increase in momentum of the electron is given by $p\left(1\right) - p\left(k^{-1}\right) = -p\left(k^{-1}\right)$. Since momentum should reverse with velocity, we demand that $p\left(k^{-1}\right) = -p\left(k\right)$, so that, in this frame too, the electrons momentum increases by $p\left(k\right)$. For momentum conservation to work in this frame, this must be decrease in the momentum of the photon. this frame is receding from the incoming photon, and hence it sees its frequency reduced by the Doppler effect to $\nu_i' = \frac{\nu_i}{k}$, while the outgoing photon gets shifted the other way, $\nu_f' = \nu_f k$.

Thus, momentum conservation leads to

$$\frac{h}{c}\left(\frac{\nu_i}{k}+\nu_f k\right)=p\left(k\right)=\frac{h}{c}\left(\nu_i+\nu_f\right)$$

and since the electron will move ($k \neq 1$), we must have

$$\nu_f=\frac{\nu_i}{k} \tag{6.7}$$

Note the power of the relativity principle, by hopping to another frame, we have already got some way to a solution, without worrying about the precise forms of $p\left(k\right)$ and $T\left(k\right)$!

With this ratio between ν_i and ν_f, it is easy to see that the kinetic energy to momentum ratio becomes

$$\frac{T}{pc}=\frac{k-1}{k+1} \tag{6.8}$$

Let's make an important point clear at this stage. The relations (6.4), (6.5) and (6.6) that I had written down so far were *equations* - they were to be solved to find the unknowns k and ν_f, given the frequency of the incident photon. However, (6.8) is an *identity* - one that must be satisfied by the kinetic energy and the momentum of the electron, no matter what its speed or k is. This is because, firstly, (6.8) expresses a relation involving the electron alone - all trace of the photon has cancelled from it! So, (6.8) is valid for whatever k factor the electron can have after the collision - but, and here is the crux of the matter, this can be any value whatsoever - all you have to do is to vary the incoming photon's frequency!

Does our classical relations for the momentum and the kinetic energy satisfy the identity (6.8)? The answer, as you can easily check, is *no*! Classically, the kinetic energy to momentum ratio is

$$\frac{T}{pc} = \frac{\frac{1}{2}mu^2}{cmu} = \frac{\beta}{2} \neq \frac{k-1}{k+1}$$

Clearly the classical definition of the momentum needs revision!

Since this is a rather shocking conclusion (although, with what you have been subjected to in the previous sections on kinematics with a flashlight, you may have been expecting such shocks!), let us take a deeper look at the argument above. Note that the classical results where holding the fort up to (6.6) - till that point there was no reason to believe that they could not fit into our scheme of things. The switch occurs when we show that *if momentum conservation is to be valid in another frame, too,* then we must have (6.7) and this leads to (6.8) which really leads to the breakdown in the classical formulae. In other words, the classical formula for the momentum can not be valid if we want its conservation to be true in all inertial frames!

One more point - in deriving (6.8) I had to assume that kinetic energy as well as momentum to be conserved. You may see a glimmer of hope there. There are inelastic collisions in classical mechanics too, where kinetic energy is not conserved. So, perhaps there is nothing wrong with our classical formulae - the failure of (6.8) is simply telling us that the electron-photon collision is, for some reason, an inelastic one? As I will show later, this possibility, too is a no go - we have no option but to revise our age old formula for the momentum!

You may be wondering why I am harping so much on the classical formula for the momentum being wrong. After all, all I have shown is that the classical ratio between the kinetic energy and the momentum of the electron is wrong. Couldn't it be that the classical momentum formula survives the rigors of STR - and it is the classical kinetic energy formula that fails? One look at (6.1) will immediately tell you that this is impossible - I can derive the formula for T from the expression for p, so $p = mu$ will always lead to $T = \frac{1}{2}mu^2$ and hence the wrong ratio. Of course, if we are forced to change the formula for p, that for T will have to change also.

Now although relativity tells us that the ratio $\frac{T}{pc}$ is $\frac{k-1}{k+1}$ and not really the classical $\frac{\beta}{2}$, it must be close to this ratio for velocities much smaller than light, *i.e.* for small values of β. You should check that this, really, is true.

> If the electron moves slowly, we may use the classical relations to solve the problem. This gives
>
> $$mu + \frac{1}{2c}mu^2 = 2h\nu_i$$
>
> (it is more convenient to use u than k for the nonrelativistic situation) giving, to the first approximation
>
> $$u \approx 2\frac{h\nu_i}{m}$$
>
> which means that
>
> $$\nu_f \approx \nu_i - 2\frac{h}{m}\nu_i^2.$$
>
> Are these value reasonable? If my flashlight emits yellow light (like reasonable flashlights) of a wavelength 600 nm, then the velocity of our electron comes out to be about 7×10^{12} m s^{-1} - many, many times the speed of light! The energy of each photon may be small - but to the electron, they are really like cannonballs!

6.4 The momentum in STR

How can we ever hope to solve the problem if we don't know either $p(k)$ or $T(k)$? The whole point is, we can use our central principle of relativity - all inertial frames are equally good, to actually deduce what these functions can be. For our first deduction, we won't have to look any further, we already have all the ingredients for deducing $p(k)$ and $T(k)$ in the identity (6.8).

All we need is the relation, obvious from (6.1) that

$$\frac{dT}{dp} = v = c\beta = c\frac{k^2-1}{k^2+1} \tag{6.9}$$

Taking the logarithm of both sides of (6.8) and differentiating gives us

$$\frac{1}{T}\frac{dT}{dp} = \left(\frac{1}{k-1} - \frac{1}{k+1}\right)\frac{dk}{dp} + \frac{1}{p}$$

which, in conjunction with (6.9) leads to

$$\frac{k^2-1}{k^2+1}\frac{k+1}{k-1}\frac{1}{p} = \frac{2}{k^2-1}\frac{dk}{dp} + \frac{1}{p}$$

and finally to

$$\frac{dp}{p} = \frac{k^2+1}{k\left(k^2-1\right)}\,dk. \tag{6.10}$$

Equation (6.10) can be integrated rather trivially to get

$$p = \frac{k^2-1}{k}A$$

where A is an arbitrary constant of integration. This result can be obviously written in the form

$$p = 2\beta\gamma A$$

and all that is left is to identify the constant A. For this, just remember that for small velocities, this formula must reduce to the classical result $mu = mc\beta$. This means that $A = \frac{1}{2}mc$ and the new relativistic formula for the momentum is

$$p = \frac{k^2 - 1}{2k} mc = mc\beta\gamma = \gamma mu = \frac{mu}{\sqrt{1 - \frac{u^2}{c^2}}}. \tag{6.11}$$

As for the kinetic energy, applying (6.8) immediately gives

$$T = \frac{(k-1)^2}{2k} mc^2 = (\gamma - 1)\, mc^2 = mc^2 \left(\frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} - 1 \right). \tag{6.12}$$

6.4.1 The speed limit

This expression for the kinetic energy helps explain why you can't accelerate a massive particle to the speed of light. Note that as the speed of the particle approaches the speed of light, its kinetic energy increases without bound. Remember that the amount by which the kinetic energy of a particle rises is equal to the work done on it[2]. So, by keeping on doing work on a massive particle you can keep on raising its kinetic energy. However, as the particle's speed gets closer and closer to the speed of light, its kinetic energy can suffer a huge increase without any appreciable gain in the speed. Since no matter how hard and long you can keep on at it, you can do only a finite amount of work implies that you can make the speed of a massive particle approach that of light - but you can never exactly attain that speed!

[2]If you are wondering whether this result, which you knew as the work-energy theorem in Newtonian mechanics, is still valid in STR, let me hasten to assure you that it is. After all, this is built-in in the definition of the kinetic energy, equation (6.1)!

What about the photon - the particle that *has* to move with the speed of light? If (6.11) and (6.12) are valid for the photon, then it would land up with infinite amounts of momentum and energy. The only way out, then, is to assume that the mass m of the photon is - zero! Since the factor γ is ∞ for a particle moving with the speed of light, the factor $m\gamma$ is not well defined - but is consistent with any finite value. The momentum carried by a photon has nothing to do with its speed - indeed, quantum mechanics tells us that it is controlled by its wavelength. One point, though - if you take the limit $m \to 0$ and $\gamma \to \infty$ in such a way that the ratio $\frac{p}{v} \to \frac{p}{c} = m\gamma$ stays finite, then the expression for the kinetic energy

$$T = mc^2 (\gamma - 1) \to \frac{p}{c} c^2 = pc$$

which is exactly the expression that we used for the photon kinetic energy, (6.3).

What about faster than light particles, *tachyons* - which used to be (and still is, perhaps to a somewhat lesser extent) the staple of science fiction writers everywher?. If you naively put in a velocity larger than c in the relativistic expressions for momentum and kinetic energy, (6.11) and (6.12), you get imaginary values - which seems to rule out faster than light travel. A better way to interpret this is simply follow what we started this section with - it takes an unbounded amount of energy to speed up a particle moving slower than light to the speed of light. Since you will not be able to speed a particle up to the speed of light, the question of going beyond does not arise!

If you think about this a bit, you will realize that there is nothing in this argument about particles that have always been moving faster than light. Such particles, on the other hand, can never be decelerated below the speed of light. Even slowing them down to the speed of light will require you to do an infinite amount of negative work! So, the upshot of it all is, from the point of view of kinetics, all that you can say is that any subluminal (slower than light) particles will have to stay subluminal, superluminal (faster than light) particles superluminal, and massless particles will have to move with the speed of light!

A much stronger argument against tachyons is the one based on causality that we saw in section 3.9. If a tachyonic particle could carry the information about the occurance of an event $\mathcal{A}$ to the location of another event $\mathcal{B}$, we saw there that it is possible for a slower than light observer to see $\mathcal{B}$ happen before $\mathcal{A}$ - thus leading to causality violation - which is a strict no-no. Of course, if tachyonic particles exist which do not interact in any way with the particles of our our subluminal world, there will be no violation of causality. However, such particles could never be detected experimentally even if they did exist - and they would not belong to the realm of physics!

6.4.2 A "better" derivation

The above derivation of the relativistic formulae for momentum and kinetic energy has one problem that you may have noticed. As I have warned before, one can, perhaps object to this on the grounds that (6.8), on which it is based relies not only on momentum conservation, but also on that of kinetic energy. I will now show you a better derivation - one that does not rely on the conservation of kinetic energy at all!

The idea behind this derivation is the same as always - if momentum conservation is to be a valid law of physics, then it is going to be equally valid in all reference frames. So far, we have applied momentum conservation to our system of a photon bouncing off an electron from two different inertial frames, one in which the electron is initially at rest and one in which it is finally at rest. Let us now write down the law of momentum conservation in an arbitrary frame, one travelling to the left with respect to our original inertial frame with a Doppler factor of K^{-1}. The reason I have chosen an observer moving to the left instead of the right is simply to make the algebra slightly simpler (you should verify that the conclusion stays unchanged if the observer were travelling to the right). Again, I have chosen to write the observer's Doppler factor as K^{-1} as opposed to K merely to ensure that K is larger than 1 - a choice that marginally helps in keeping track of the terms.

The whole point behind using the Doppler factor instead of the speed should become clear now. In section 3.4, I showed you that the velocity addition formula is much simpler in terms of the former, rather than the latter. In our new frame, the electron initially has a Doppler factor of K, while after collision its Doppler factor becomes kK. What about the photon. The observer is approaching the incident photon and hence she sees it blue-shifted to a frequency $\nu_i K$, while the final photon is red-shifted to $\frac{\nu_f}{K}$. Thus, in this new frame, the law of conservation of momentum becomes

$$p\left(kK\right)-p\left(K\right)=\frac{h}{c}\left(K\nu_i+\frac{\nu_f}{K}\right)$$

Using (6.4) and (6.7), this becomes

$$\frac{p\left(kK\right)-p\left(K\right)}{p\left(k\right)}=\frac{K^2k+1}{K\left(k+1\right)} \tag{6.13}$$

A moments though will tell you that our function $p\left(k\right)$ must satisfy (6.13) not only for all values of K, but also for all values of k! The task, then, is to deduce the functional form for a function satisfying this equation.

While solving functional equations is quite a difficult task in general, there is a simple trick that helps us to solve (6.13). The right hand side of equation (6.13) becomes simply $\frac{K^2+1}{2K}$ when $k = 1$, but the left hand side, for this value of k is of the form $\frac{0}{0}$. You can, of course find the limit of the left hand side as $k \to 1$, which is simply $\frac{Kp'(K)}{p'(1)}$ from l'Hospital's rule. Thus

$$p'(K) = \frac{K^2+1}{2K^2}p'(1) \tag{6.14}$$

Since $p'(1)$ is a constant, this is easily solved to yield the formula, (6.11) that we have already found out before for the momentum.

Once I have shown you what the momentum is, you can simply use (6.8) to find out the kinetic energy. In case you have qualms about using (6.8) at all, you are welcome to carry out the integration in (6.1) to check that you arrive at the same expression for the kinetic energy as (6.12).

6.5 Mass is energy

The expression for kinetic energy of a point particle is, then, not $\frac{1}{2}mu^2$ but $(\gamma - 1)\,mc^2$. The two expressions match, as they must, when the velocity is small and of course they both vanish for a stationary particle. Now, the fact that the kinetic energy vanishes for a stationary particle makes eminent sense - it is meant to be the energy that a body has due to its *motion*, after all! Remember, though, that only energy differences have a physical meaning and not actual values of the energy. So, you are always free to add any constant to the energy without worrying about the physical consequences. Given the choice, what would you think is the most natural constant to add to the expression (6.12) for the kinetic energy? I am sure most of you would like to add the constant mc^2 so that the right hand side becomes a lot neater! Of course, this means that the energy of a stationary particle is no longer zero but mc^2 - so calling it the kinetic energy is not very justified. Physicists call this quantity the total energy E, and hence expression (6.12) becomes

$$E = T + mc^2 = (\gamma - 1)\, mc^2 + mc^2 = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{6.15}$$

an equation which you must all know as the one that has made Einstein a household name[3]!

There you have it - $E = \gamma mc^2$- arguably the most famous equation in physics! Yet, I am sure you are feeling a bit disappointed with the way in which I introduced it. "Is it true, then -" you must be asking yourselves, "that physics' most famous equation is a bit of a sham - simply an attempt to window dress an expression for kinetic energy so that it looks a little better?" If all collisions in the world were elastic, then all this new expression would have been is a mere rewrite of (6.12). Take heart, though, in the case of processes which the kinetic energy does not balance out, equation (6.15) takes a new, deeper significance - one that we will now explore.

Let's start by considering the most simple example of an inelastic collision that I can think of - two particles of identical mass $m$colliding head on with equal and opposite velocities, corresponding to Doppler factors of k and k^{-1}, respectively, and sticking together. You know that after the collision, the composite particle will be standing still. From your high school education, you would

[3]Perhaps you are more used to hearing $E = mc^2$ instead of $E = \gamma mc^2$. Wait till section 6.9 to see how these two expressions are really one and the same!

conclude that the final particle has a mass of $2m$, and the total kinetic energy of $\frac{1}{2}mu^2 \times 2 = mu^2$ that the incident pair had will be converted to heat energy that will, at least initially, heat up the composite body. Let's see what our new-found laws of relativistic mechanics tell us about this scenario.

To feel the real impact of STR on this simple problem, let me now describe the same collision from the point of view of another inertial frame, moving with a Doppler factor of K^{-1} with respect to our first one. In this frame, the two colliding particles have Doppler factors of kK and $k^{-1}K$, respectively and the final, composite particle one of K. Now, the net initial momentum of the system works out to be

$$\begin{aligned} p_i &= p\left(kK\right) + p\left(\frac{K}{k}\right) \\ &= \frac{1}{2}mc\left(kK - \frac{1}{kK}\right) + \frac{1}{2}mc\left(\frac{K}{k} - \frac{k}{K}\right) \\ &= \frac{mc}{2kK}\left(k^2K^2 - 1 + K^2 - k^2\right) \\ &= \frac{mc}{2kK}\left(K^2 - 1\right)\left(k^2 + 1\right) \\ &= \frac{1}{2}mc\left(K - \frac{1}{K}\right)\left(k + \frac{1}{k}\right) \\ &= \frac{1}{2}Mc\left(K - \frac{1}{K}\right) \end{aligned}$$

where I have chosen to write the combination $m\left(k + \frac{1}{k}\right)$ as M. Since the law of conservation of momentum, if it is to be a valid physical law, must be valid in all inertial frames, this must be the momentum of the composite particle in the new frame. But the Doppler factor of the composite particle in this frame being K, the mass of the composite particle must be M! Note that though we used a new frame to find out the mass of the composite particle this value is independent of the frame of reference. Thus the final particle has a mass of, not $2m$, but

$$M = 2m \times \frac{1}{2}\left(k + \frac{1}{k}\right) = 2m\gamma \tag{6.16}$$

which is larger than the net initial mass by $2(\gamma - 1)m$. The conservation of mass, one of the oldest laws that we had learned in the physical sciences is not valid in relativity!

In our first frame, the net kinetic energy of the incident particles was $2(\gamma - 1)mc^2$, while that of the final particle was, of course, zero. Thus, what you loose in kinetic energy is exactly what you gain in mass, once you take the factor of c^2 into account. You may be suspicious that this nice result may be valid only in the initial frame, so let us try out the calculation in the new arbitrary frame.

$$\begin{aligned}
T_i &= T(kK) + T\left(\frac{K}{k}\right) \\
&= \frac{1}{2}mc^2\left(kK + \frac{1}{kK}\right) - mc^2 + \frac{1}{2}mc^2\left(\frac{K}{k} + \frac{k}{K}\right) - mc^2 \\
&= \frac{mc^2}{2kK}\left(k^2K^2 + 1 + K^2 + k^2\right) - 2mc^2 \\
&= \frac{1}{2}Mc^2\left(K + \frac{1}{K}\right) - 2mc^2
\end{aligned}$$

where I have used (6.16). Now, the first term on the left hand side above is the total energy $E_f = T_f + Mc^2$ of the final particle, and thus

$$T_i = T_f + (M - 2m)c^2$$

and this means that even in this frame, the increase in mass is

$$M - 2m = \frac{1}{c^2}(T_i - T_f) \tag{6.17}$$

- *i.e.*, the decrease in the kinetic energy, divided by c^2.

Another way of writing the equation above immediately suggests itself. Equation (6.17) can be easily rewritten into

$$T_i + 2mc^2 = T_f + Mc^2$$

which means that

$$E_i = E_f.$$

So, the quantity E that we defined by adding mc^2 to the kinetic energy has a great significance of its own. In this inelastic two body process, it is E which is conserved - not T or the total mass! This is a big departure from classical physics - in classical physics the conservation of energy and that of mass were two separate laws - here they are rolled up into one common conservation law! Note that all that we needed to show this is that momentum conservation is a valid physical law (and hence equally valid in all inertial frames). The only thing that we need to check now is whether this is a one-off result - valid only for two particles colliding and sticking together, or is this a much more general law of physics.

6.6 The conservation of energy from that of momentum

Consider a general process where the initial state has particles of mass $m_1, m_2, \ldots, m_r$ which are moving with Doppler factors $k_1, k_2, \ldots, k_r$, respectively. In the final state, the number and masses of particles may be different[4] . Let the masses and Doppler factors be $\mu_1, \mu_2, \ldots, \mu_s$ and $\kappa_1, \kappa_2, \ldots, \kappa_s$ in the final state.

[4]Imagine the decay of the neutron. The initial state has one particle only - the neutron. The final state has three - the proton, the electron and the antineutrino.

The law of conservation of momentum in this frame is, then

$$\frac{1}{2}m_1c\left(k_1-\frac{1}{k_1}\right)+\frac{1}{2}m_2c\left(k_2-\frac{1}{k_2}\right)+\ldots+\frac{1}{2}m_rc\left(k_r-\frac{1}{k_r}\right)=$$
$$\frac{1}{2}\mu_1c\left(\kappa_1-\frac{1}{\kappa_1}\right)+\frac{1}{2}\mu_2c\left(\kappa_2-\frac{1}{\kappa_2}\right)+\ldots+\frac{1}{2}\mu_sc\left(\kappa_s-\frac{1}{\kappa_s}\right) \quad (6.18)$$

Allow me to bring in our old friend - the observer moving with a Doppler factor K^{-1}. She sees the initial Doppler factors as $k_1K, k_2K, \ldots, k_rK$ and the final ones as $\kappa_1K, \kappa_2K, \ldots, \kappa_sK$. If momentum is to be conserved for her, too, we have

$$\frac{1}{2}m_1c\left(k_1K-\frac{1}{k_1K}\right)+\ldots+\frac{1}{2}m_rc\left(k_rK-\frac{1}{k_rK}\right)=$$
$$\frac{1}{2}\mu_1c\left(\kappa_1K-\frac{1}{\kappa_1K}\right)+\ldots+\frac{1}{2}\mu_sc\left(\kappa_sK-\frac{1}{\kappa_sK}\right) \quad (6.19)$$

which can be easily rearranged to give

$$\frac{c}{2}\left[\left(m_1k_1+m_2k_2+\ldots+m_rk_r\right)K-\left(\frac{m_1}{k_1}+\frac{m_2}{k_2}+\ldots+\frac{m_r}{k_r}\right)\frac{1}{K}\right]=$$
$$\frac{c}{2}\left[\left(\mu_1\kappa_1+\mu_2\kappa_2+\ldots+\mu_s\kappa_s\right)K-\left(\frac{\mu_1}{\kappa_1}+\frac{\mu_2}{\kappa_2}+\ldots+\frac{\mu_s}{\kappa_s}\right)\frac{1}{K}\right] \quad (6.20)$$

Since this is to be equally true for all values of K, we must have

$$m_1k_1+m_2k_2+\ldots+m_rk_r = \mu_1\kappa_1+\mu_2\kappa_2+\ldots+\mu_s\kappa_s \quad (6.21)$$

and

$$\frac{m_1}{k_1}+\frac{m_2}{k_2}+\ldots+\frac{m_r}{k_r} = \frac{\mu_1}{\kappa_1}+\frac{\mu_2}{\kappa_2}+\ldots+\frac{\mu_s}{\kappa_s} \quad (6.22)$$

- a rather striking result!

One immediate consequence of (6.21) and (6.22) is

$$\frac{c^2}{2}\left[\left(m_1k_1+m_2k_2+\ldots+m_rk_r\right)+\left(\frac{m_1}{k_1}+\frac{m_2}{k_2}+\ldots+\frac{m_r}{k_r}\right)\right]=$$

$$\frac{c^2}{2}\left[(\mu_1\kappa_1+\mu_2\kappa_2+\ldots+\mu_s\kappa_s)+\left(\frac{\mu_1}{\kappa_1}+\frac{\mu_2}{\kappa_2}+\ldots+\frac{\mu_s}{\kappa_s}\right)\right] \tag{6.23}$$

which is nothing but the equation of conservation of energy,

$$m_1c^2\left(\frac{k_1^2+1}{2k_1}\right)+\ldots+m_rc^2\left(\frac{k_r^2+1}{2k_r}\right)=\mu_1c^2\left(\frac{\kappa_1^2+1}{2\kappa_1}\right)+\ldots+\mu_sc^2\left(\frac{\kappa_s^2+1}{2\kappa_s}\right) \tag{6.24}$$

Although I have written down this equation for the observer I started out with, it is rather trivial to show that this conservation will work for any arbitrary observer (show it!).

Thus, demanding that the law of conservation of momentum is equally valid for all inertial observers leads to the law of conservation of energy as well, for a rather general process. You may have already caught on to this - this works just as well the other way round, too. Demanding that energy is conserved in all frames leads to momentum conservation. Surely, then, there must be a deeper connection between momentum and energy in relativity? There is! I will discuss this connection in the very next section.

6.7 The momentum four vector

So far, I have made use of the fact that any law of physics must stay intact when you jump from one frame of reference to another to deduce the form that momentum and energy must have. Let us now take a look at how the relativistic momentum and energy do transform under a Lorentz transformation.

Alice observes a particle to have the Doppler factor k. We have already seen that this particle must have a momentum and energy given by

$$
\begin{aligned}
p\left(k\right) &= \frac{mc}{2}\left(k-\frac{1}{k}\right)\\
E\left(k\right) &= \frac{mc^{2}}{2}\left(k+\frac{1}{k}\right)
\end{aligned}
$$

which yields, rather trivially

$$k = \frac{1}{mc}\left(\frac{E}{c}+p\right) \tag{6.25-a}$$

$$\frac{1}{k} = \frac{1}{mc}\left(\frac{E}{c}-p\right) \tag{6.25-b}$$

One result that follows immediately by multiplting the two equations above is

$$\frac{E^{2}}{c^{2}}-p^{2}=m^{2}c^{2}$$

Although I have proven this for a 1 dimensional world, te result is perfectly valid in three space dimensions too - as we will see in a while.

If Bob moves away from Alice with a Doppler factor K, he will see the particle to have a Doppler factor of $\frac{k}{K}$. So, according to Bob, the particle has momentum and energy given by

$$
\begin{aligned}
p'\left(k\right) &= \frac{mc}{2}\left(\frac{k}{K}-\frac{K}{k}\right)\\
E'\left(k\right) &= \frac{mc^{2}}{2}\left(\frac{k}{K}+\frac{K}{k}\right)
\end{aligned}
$$

You can use (6.25-a) and (6.25-b) to rewrite Bob's measured values in terms of those measured by alice. This leads to

$$p' = \frac{1}{2}\left(K+\frac{1}{K}\right)p-\frac{1}{2}\left(K-\frac{1}{K}\right)\frac{E}{c}$$

$$\frac{E'}{c} = -\frac{1}{2}\left(K - \frac{1}{K}\right)p + \frac{1}{2}\left(K + \frac{1}{K}\right)\frac{E}{c}$$

and using our by now familiar values for the quaantities $\frac{1}{2}\left(K + \frac{1}{K}\right)$ and $\frac{1}{2}\left(K - \frac{1}{K}\right)$ leads to

$$p' = \gamma\left(p - \beta\frac{E}{c}\right) \tag{6.26-a}$$

$$\frac{E'}{c} = \gamma\left(-\beta p + \frac{E}{c}\right) \tag{6.26-b}$$

Do the above equation look familiar? They should - they are exactly the same as the transformations of x and ct! The all important point right now, though, is that the value of the momentum in Bob's frame depends not only on the value of the momentum that Alice measures - but also on the value of the energy according to her. Thus, if only momentum is conserved according to Alice and not the energy - then even the momentum will not be conserved according to Bob. Thus, we can ultimately trace the coonection between the law of conservation of momentum to that of energy to the fact that as far as the transformation from one inertial observer to another is concerned, the pair $\left(p, \frac{E}{c}\right)$ behaves just like the pair (x, ct).

To those of you who have not skipped the section on vectors and scalars in spacetime, this should not come as a big surprise (If you were the impatient sort, now is a good time to go back and read section 5.5). After all, we have seen that four-vectors behave just like spacetime coordinates under Lorentz transformations. So, it seems that the pair $\left(p, \frac{E}{c}\right)$ that we have been seeing so far is just a part of some four-vector.

In Newtonian physics momentum is three-vector - simply because it is a product of a scalar - the mass, and a vector - the velocity. The analogous four dimensional quantity is simple - all you have to do to define the momentum four-vector is multiply mass m by the four dimensional generalisation of the vector $\vec{u}$, which is none other than the four-velocity that we met in subsection 5.5.4. Thus, we have the four-momentum

$$p^{\mu} \equiv mU^{\mu} = m\gamma\left(c, \vec{u}\right) \tag{6.27}$$

Note that the spatial part of this four-vector is precisely $\gamma m\vec{u}$ - the formula that we have found out a while ago for the accurate relativistic form of the momentum (albeit in one dimensional form). As an added bonus note that the temporal part of p^{μ} is γmc which is just $\frac{E}{c}$ - so that in one swoop we have found out the relativistic versions of both the momentum and the energy!

In fact now you see why adding the rest energy mc^2 to the kinetic energy to get the total energy $E = \gamma mc^2$ was such a good idea! Quiet unwittingly, we had stumbled on the temporal part of the four-vector whose spatial part is the momentum three-vector. This also tells us why the demand that the conservation of momentum be valid in all frames immediately leads to the conservation of energy as an added bonus. It is actually the four-momentum that is conserved - the conservation of momentum and that of energy are two pieces of this single conservation law!

One invariant quantity that can be built immediately, given a four-vector, is its norm. Finding out the norm of the energy-momentum four-vector is really trivial - because we already know the norm of the four-velocity, U^{μ}. So,

$$P \cdot P = P^{\mu}P_{\mu} = m^2U^{\mu}U_{\mu} = m^2c^2 \tag{6.28}$$

- which is trivially an invariant. If you write out the norm of the energy-momentum four-vector explicitly, this gives a very useful identity

$$\frac{E^2}{c^2} - \vec{p} \cdot \vec{p} = m^2 c^2$$

- a formula that we have seen a while ago in which is often written in the form

$$E^2 = p^2 c^2 + m^2 c^4 \quad (6.29)$$

often informally called, for obvious reasons, the relativistic Pythagoras law.

Another invariant that is very often quite useful is the scalar product of two four-momenta P_1 and P_2. A direct calculation yields

$$P_1 \cdot P_2 = m_1 m_2 \gamma(u_1) \gamma(u_2) \left(c^2 - \vec{u}_1 \cdot \vec{u}_2\right)$$

However, a much more useful version of this invariant quantity can be obtained if you jump over to the rest frame of one of the particles - say the first one. Then, the two four-momenta are $m_1 \left(c, \vec{0}\right)$ and $m_2 \gamma(v_{\text{rel}}) (c, \vec{v}_{\text{rel}})$, respectively, and the invariant product takes the form

$$P_1 \cdot P_2 = m_1 m_2 \gamma(v_{\text{rel}}) c^2 \quad (6.30)$$

Of course, this is identically equal to the expression that I found by the direct calculation - all you have to do to realize this is look at equation (5.24).

A direct consequence of the above equation concerns the norm of the net four-momentum of a system of particles

$$P = P_1 + P_2 + \ldots + P_n$$

Since the γ factor for any velocity obeys $\gamma(v) \geq 1$ (with the equality holding only for $v = 0$) we get

$$\begin{aligned} P^2 &= P_1^2 + \ldots + P_n^2 + 2P_1 \cdot P_2 + 2P_1 \cdot P_3 + \ldots + 2P_{n-1} \cdot P_n \\ &= c^2 \left(m_1^2 + \ldots + m_n^2 + 2m_1 m_2 \gamma \left(v_{12} \right) + \ldots + 2m_{n-1} m_n \gamma \left(v_{n-1,n} \right) \right) \\ &\geq c^2 \left(m_1^2 + \ldots + m_n^2 + 2m_1 m_2 + \ldots + 2m_{n-1} m_n \right) \\ &= c^2 \left(m_1 + \ldots + m_n \right)^2 = c^2 M^2 \end{aligned}$$

where $M = m_1 + \ldots + m_n$ is the total mass of the system. Thus

$$P^2 \geq c^2 M^2 \tag{6.31}$$

with equality holding only when all the particles are moving together and thus all the inter-particle relative velocities vanish.

This is as good a point as any to pay attention to a special case - the massless particle. We have already seen one example - the photon. As we have already seen, massless particles have to travel at the speed of light and for them our energy-momentum formulae need to be applied with some care. We have already seen that in the massless limit

$$T = pc.$$

Now, for a photon, there is no rest energy - which means that $E = T$. So, the photon obeys

$$E = pc \tag{6.32}$$

which, as you can easily verify, is what (6.29) reduces to for $m = 0$. For a massless particle, the energy-momentum four-vector reduces to

$$P^\mu \equiv (p, \vec{p}) \tag{6.33}$$

where $p = |\vec{p}|$. Fore a photon, quantum theory gives us the momentum -wavevector relation

$$\vec{p} = \hbar\vec{k} \tag{6.34}$$

leading to the form

$$P^{\mu}_{\gamma} \equiv \hbar\left(k, \vec{k}\right) \tag{6.35}$$

One important property of the energy-moemntum four-vector for massless particles is obvious from (6.28) - the norm of such a four vector is

$$P^2 \equiv P^{\mu}P_{\mu} = 0. \tag{6.36}$$

What happens to our expression for $P_1 \cdot P_2$ if one of the two particles is a photon. Of course, you can jump over to the frame where the other particle is at rest and this immediately gives

$$P_1 \cdot P_2 = \hbar k_0 m \tag{6.37}$$

where k_0 is the wavenumber of the photon in the rest frame of the other particle. This shortcut, of course, is not available when both particles are photons - you can't step into a photon's "rest frame"! In this case, we have to use the explicit expression

$$P_1 \cdot P_2 = \hbar^2\left(k_1k_2 - \vec{k}_1 \cdot \vec{k}_2\right). \tag{6.38}$$

6.8 Covariance of the conservation of energy-momentum

The fact that the four-momentum is a *bona fide* four-vector may lead you to think that this makes the conservation of momentum manifestly covariant (as discussed in subsection 5.5.7) - and thus I can avoid the issue of checking whether it stays equally valid in all frames. There is a slightly subtle point to be wary of here, though! It is true that the four-momentum of a single particle is a four-vector, but is the total four-momentum of a system (after all, this is what is conserved in a process) one? There seems to be no doubt about this - the sum of two four-vectors is one, and this can be easily extended to the sum of many four vectors! The subtlity comes in from the fact that when you change from one inertial observer to another, you are not really summing up the *same* four-momenta to get the total four-momentum! The point is, the total four-momentum *is* the sum of all the four-momenta of the particles in the system, but it has to be the sum of all the four-momenta at a given instant. Jumping to another frame changes the meaning of "a given instant" - which events are simultaneous and which are not depends on the observer! So, the initial total four-momentum P_i (which is equal to the final total momentum P_f) goes over to P_i' (which must equal P_f') upon a switch of observers - but the P_i' is not the same as the Lorentz-transformed P_i, neither is P_f' the same as the Lorentz-transformed P_f. If we denote the Lorentz-transformed versions of P_i and P_f by $\overline{P}_i'$ and $\overline{P}_f'$, respectively, it is obvious that $P_i = P_f$, which is the law of conservation of moemntum for the first observer, will imply $\overline{P}_i' = \overline{P}_f'$. Unfortunately, this is *not* the law of conservation of momentum for the second observer, which is $P_i' = P_f'$.

Why, then, do we believe in the conservation of four-momentum? To explain this, I will take the help of a diagram, figure .

6.9 Do we need a relativistic mass?

You may be a bit confused with the equations I have written for relativistic mechanics. Maybe the one that you will have the most difficulty in digesting is $E = \gamma mc^2$- you must have heard from childhood that Einstein's famous formula for mass-energy equivalence is $E = mc^2$ - so where did this extra factor of γ come from?

To see the reason behind this difference in what you always knew and what I am claiming to be the correct expression we will have to go back a bit - to the point where I introduced the correct relativistic formula for the momentum.

$$p = \gamma mu = \frac{mu}{\sqrt{1 - \frac{u^2}{c^2}}}$$

At this point there are two possible courses of action open before us. The first one, the one I have followed so far, is to gracefully accept that the formula for momentum is not the old one of mass times velocity, it is a different expression that, however, does reduce to the old formula in the case of slowly moving particles.

The other option is one that was very popular in the early days of relativity. This one insisted that the momentum is still the same old mass times velocity - only the mass is no longer the good old constant quantity that we had been used to in the early days, but one that increases with speed according to

$$m = \frac{m_0}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{6.39}$$

Of course, all that has been done here is absorb the factor γ in m to define this new quantity. In this second way of thinking about things, the left hand side of the above equation is called the *relativistic* mass - while m_0, which is what I have been calling the mass m all this time, is the value of this relativistic mass at $u = 0$, and is hence called the *rest mass*. This explains the mystery of the missing γ in the arguably most famous equation of all physics - it is simply that the m in $E = mc^2$ is the reativistic mass (our $m\gamma$).

Since bringing the relativistic mass leads to simplification of two of the major equations of relativistic mechanics - the definition of the momentum and the formula for mass-energy equivalence - it seems like an eminently good idea.The trouble here is that hiding the factor γ in the definition of the mass only tends to conceal the basic difference between Newtonian and Einsteinian kinetics - the fact that the definition of momentum has changed! Perhaps more importantly, it tends to make you think that all you have to do is to replace the rest mass in each Newtonian formula by the relativistic mass and you are done! That is not the case after all. Indeed perhaps the best known equation of Newtonian mechanics

$$\vec{F} = m\vec{a}$$

does not generalize nicely at all to a new relativistic formula. I will have a lot more to say about this in section 6.10. For the time being, let me just say that in the context of the generalized second law, we can talk of (at least) two kinds of mass - longitudinal mass and transverse mass! Moreover, the transverse mass happens to be numerically the same as the relativistic mass. Given all this plethora of masses - the utility of the relativistic mass concept loses some of its sheen.

Today, most people are of the opinion that the relativistic mass is too much of a baggage to carry around for the little cosmetic benefits it offers. What I have done here is conform to this trend. Except in the equation (6.39) above, all my m's denote the mass of a particle - the parameter that once used to be called the rest mass.

There is, perhaps, a sociological lesson here. When STR came into the scene - it brought with it the hope (or threat, depending on your point of view) of sweeping all the old concepts away. At that stage, the most striking feature of the theory was that certain things that we had always held to be absolute actually turns out to be observer dependent! This, after all, was summed up by the catchphrase - "everything is relative!" At that point, the fact that mass was seen as dependent on velocity is just one more addition to the long list of "relative" entities. Today, perhaps, the emphasis has shifted. We regard the theory of relativity today, not so much as a theory of what is relative (observer dependent) but as a theory in which the emphasis is on the fact that the laws of physics (all of them) are observer independent! So much so, that many people has strongly advocated a change in the name of the theory from the "theory of relativity" to the "theory of invariance"! In the modern way of looking at things, the stress is on quantities that do not depend on the observer. The so called rest mass (also known as the proper mass) of a particle is such a quantity. That is, to my mind, the single most important reason for giving the (rest) mass its rightful status as *the* mass.

6.10 Force and acceleration in STR

Now that we have been forced to accept a new definition of the momentum in order to ensure that the law of conservation of momentum be Lorentz covariant, the natural question is - what happens to the central equation of Newtonian kinetics - the second law? Remarkably, the second law stays intact - not in the form

$$\vec{F} = m\vec{a} \tag{6.40}$$

that may be familiar to you, but in the form originally expressed by Newton

$$\vec{F} = \frac{d\vec{p}}{dt}. \tag{6.41}$$

The only difference is, here we have to use the new definition of the momentum, $\vec{p} = \gamma m\vec{u}$ instead of our old formula $\vec{p} = m\vec{u}$. Since the factor γ depends on $\vec{u}$, and through it, is time dependent, it is natural that the 2nd law will take a more complicated form in terms of the acceleration than the simple Newtonian form (6.40).

The precise relation between the force and the acceleration is easy to find out from (6.41).

$$\vec{F} = \frac{d}{dt}\left(m\gamma\vec{u}\right) = m\gamma\frac{d\vec{u}}{dt} + \frac{d}{dt}\left(m\gamma\right)\vec{u} = m\gamma\vec{a} + \frac{1}{c^2}\frac{dE}{dt}\vec{u}$$

Now, one would expect that

$$\frac{dE}{dt} = \frac{dK}{dt} = \nabla_{\vec{p}}K \cdot \frac{d\vec{p}}{dt} = \vec{u} \cdot \vec{F}, \tag{6.42}$$

which is (perhaps surprisingly) the same as the formula for power that we had learnt in classical physics. As it turns out, though, this formula does not really work in all situations! It is easy to see that it will not work, if the mass (which, remember, is what many other people call the rest mass), instead of being a constant time-independent quantity, changes with time, because then you would have to take the rate of change of the rest energy into account, too! For the time being, we will confine ourselves to the case where m is a constant.

Using the above expression for the power, we can easily write

$$\vec{F} = m\gamma\vec{a} + \frac{1}{c^2}\left(\vec{F} \cdot \vec{u}\right)\vec{u} \tag{6.43}$$

which is our relativistic replacement for Newton's second law (at least for bodies with constant rest mass). As you can easily see, the presence of the second term says that the acceleration of a moving body is, in general, not in the same direction as the force. There are only two cases where the force and the acceleration are co-directional - one, when the second term vanishes because the force is perpendicular to the velocity, and two, when the force is parallel to the velocity. In the first case, we have

$$\vec{F} = m\gamma\vec{a}$$

while in the second,

$$\vec{F} = m\gamma\vec{a} + \frac{u^2}{c^2}\vec{F}$$

and hence

$$\vec{F} = m\gamma^3\vec{a}$$

This means that in the two cases where the relativistic relation between the force and the acceleration mimics the classical one, the ratio, which you would be inclined to call the mass, takes different values. So, as far as generalizing $\vec{F} = m\vec{a}$ is concerened, we have two kinds of mass - the *transverse mass* m_t, which happens to be equal to the so called relativistic mass $m\gamma$, and the *longitudinal mass* m_l which is bigger by a further factor of γ^2.

You may be slightly surprised at my assertion that (6.41) is the correct generalization of Newton's second law to the relativistic case. After all, $\frac{d\vec{p}}{dt}$ can not be the spatial part of a four-vector, for the same reason that $\vec{v}$ is not! Surely it would have been more sensible to have

$$\vec{F}_M = \frac{d\vec{p}}{d\tau} = \gamma\vec{F} \tag{6.44}$$

as the relativistic force (the M stands for Minkowski)? One way to answer this is simply to say that both Newton's second law and its relativistic versions are only half-laws! By itself Newton's second law gets reduced to merely the definition of what we call the force - the only way it gets physical content is via teaming up with another half-law - one which expresses the force in terms independent of the acceleration[5]. Thus, from one point of view, which one of the two ($\vec{F}$ and $\vec{F}_M$) to accept as the force, is a matter of convention - you can always adjust the extra γ in the half-law that helps you calculate the force!

Having said that, our vote still goes in favour of (6.41). One reason for this is, as we will see in chapter 7, this will help us to leave the Lorentz force law for the force acting on a charged particle moving in an electromagnetic field intact (no need for an extra γ factor). Another reason is that we have already used the work done by this force to deduce the relativistic expression for the kinetic energy - and this has already given us the convenient notion of the total energy that forms a nice four-vector along with the momentum.

Finally, we will be able to recover Newton's third law, at least for collisions[6], with this version of the force! When two particles collide, they stick together for a very short period of time -bduring which they exert forces on each other.

---

[5]Of course, this is true only for non-constraint forces. Constraint forces, on the other hand, do adjust themselves to ensure that the constraints are maintained - depending on the acceleration of the particle if necessary (remember the "loss of weight" of a body in a downward accelerating lift?)!

[6]For forces acting between two particles at a distance, there is no chance of getting something akin to the third law - simply because two observers will never agree on simultaneity when the two bodies are separated!

6.11 The force and acceleration four vectors

As we have already seen, the basic 4-vector is the position 4-vector

$$X = (ct, \vec{r})$$

We have already met two other 4-vectors that follow very simply from it

$$\begin{aligned} U &\equiv \frac{dX}{d\tau} = \frac{dt}{d\tau}\frac{dX}{dt} = \gamma(u)(c, \vec{u}) \\ P &\equiv mU = m\gamma(u)(c, \vec{u}) = \left(\frac{E}{c}, \vec{p}\right) \end{aligned}$$

Of course, we can (and will) define the force four-vector as

$$F = \frac{dP}{d\tau} = \gamma(u)\left(\frac{1}{c}\frac{dE}{dt}, \frac{d\vec{p}}{dt}\right) = \left(\frac{\gamma(u)}{c}\frac{dE}{dt}, \gamma(u)\vec{F}\right)$$

where $\vec{F}$ is our old friend - the 3-force. Defining $\gamma(u)\vec{F}$ as the *Minkowski force* $\vec{F}_M$ allows us to write

$$F = \left(\frac{\gamma(u)}{c}\frac{dE}{dt}, \vec{F}_M\right).$$

At this point we may feel sorely tempted to replace the $\frac{dE}{dt}$ above by $\vec{F}\cdot\vec{v}$. However, as I already hinted in the last section, this formula is not valid in general - but only in the special case where m is a constant. To see what the general result is, we will start from the simple relation

$$P^2 = P \cdot P = m^2c^2$$

and differentiate both sides with respect to τ to get

$$\frac{d}{d\tau}P^2 = 2P \cdot \frac{dP}{d\tau} = 2P \cdot F = 2mc^2\frac{dm}{d\tau} = 2mc^2\gamma\left(u\right)\frac{dm}{dt}$$

and thus

$$P \cdot F = mc^2\gamma\left(u\right)\frac{dm}{dt}$$

Using $P = m\gamma\left(u\right)\left(c, \vec{u}\right)$ and $F = \gamma\left(u\right)\left(\frac{1}{c}\frac{dE}{dt}, \vec{F}\right)$ yields, however

$$P \cdot F = m\left[\gamma\left(u\right)\right]^2\left(\frac{dE}{dt} - \vec{F} \cdot \vec{u}\right)$$

Equating these two expressions for $P \cdot F$ yields

$$\frac{dE}{dt} = \vec{F} \cdot \vec{u} + \frac{c^2}{\gamma\left(u\right)}\frac{dm}{dt} \tag{6.45}$$

which is the complete expression for power in relativity. As you can see, this differs from the classical expression due to the presence of the term involving the rate of change of the rest mass.

Using $\frac{dE}{dt} = \vec{F} \cdot \vec{u}$ the four-force can be written as

$$F = \gamma\left(\frac{1}{c}\left(\vec{F} \cdot \vec{u}\right), \vec{F}\right) = \left(\frac{1}{c}\left(\vec{F}_M \cdot \vec{u}\right), \vec{F}_M\right) \tag{6.46}$$

which of course, is valid only for the case where m is a constant. Of course, the four-force does obey a relation that is very reminiscent of Newton's second law

$$F = \frac{dP}{d\tau} = \frac{d}{d\tau}\left(mU\right) = m\frac{dU}{d\tau} = mA \tag{6.47}$$

This formula, once again, is only valid for constant rest mass m.

(6.47) gives us another way of deriving (6.43). Using our expression for the four-acceleration (5.25) and equating each component on two sides of (6.47)

$$\begin{aligned} \frac{\gamma}{c}\left(\vec{F}\cdot\vec{u}\right) &= \frac{m\gamma^4}{c}\vec{u}\cdot\vec{a} \\ \gamma\vec{F} &= m\gamma^2\vec{a} + \frac{m\gamma^4}{c^2}\left(\vec{u}\cdot\vec{a}\right)\vec{u} \end{aligned}$$

The first of these equations tells us that $\vec{F}\cdot\vec{u} = m\gamma^3\left(\vec{u}\cdot\vec{a}\right)$, a result that you could have also derived by taking the dot product of both sides of (6.43) with the velocity u. using this reult in the second equation immediately gives us (6.43).

One question that will be of some importance is the law of transformation of forces. Since our force three-vector is not the spatial part of the force four-vector, you should expect it to have a more complicated transformation law then that of, say, the three-momentum. However, though $\vec{F}$ is not the spatial part of a 4-vector, $\gamma(u)\vec{F}$ is. Remember, we saw the same thing happen with $\vec{u}$ - it is $\gamma(u)\vec{u}$ which forms the spatial part of a 4-vector, not $\vec{u}$ itself. Thus, the 3-force should transform in a way that is similar to the 3-velocity!

Let's work out the transformation equation for the 3-force for a special Lorentz transformation. The spatial part of the position 4-vector transforms under this as

$$\begin{aligned} x^{1\prime} &= \gamma\left(x^1 - \beta x^0\right) \\ x^{2\prime} &= x^2 \\ x^{3\prime} &= x^3 \end{aligned}$$

It is easy to write down the transformations for the spatial part of the 4-vector F by comparison

$$\begin{aligned}
\gamma\left(u'\right)F^{1\prime} &= \gamma\left(\gamma\left(u\right)F^{1}-\beta\frac{\gamma\left(u\right)}{c}\frac{dE}{dt}\right)=\gamma\gamma\left(u\right)\left(F^{2}-\frac{v}{c^{2}}\frac{dE}{dt}\right)\\
\gamma\left(u'\right)F^{2\prime} &= \gamma\left(u\right)F^{2}\\
\gamma\left(u'\right)F^{3\prime} &= \gamma\left(u\right)F^{3}
\end{aligned}$$

Dividing both sides by the result $\gamma\left(u'\right)=\gamma\gamma\left(u\right)\left(1-u_{1}v/c^{2}\right)$ (which follows from the transformation of the 0-th component of the 4-velocity) yields the force transformationsF_{2}\sqrt{1-\frac{v^{2}}{c^{2}}}

$$\begin{aligned}
F^{1\prime} &= \frac{F^{1}-\frac{v}{c^{2}}\frac{dE}{dt}}{1-\frac{u_{1}v}{c^{2}}}\qquad\left(=\frac{F^{1}-\frac{v}{c^{2}}\left(\vec{F}\cdot\vec{u}\right)}{1-\frac{u_{1}v}{c^{2}}}\right)\\
F^{2\prime} &= \frac{F^{2}\sqrt{1-\frac{v^{2}}{c^{2}}}}{1-\frac{u_{1}v}{c^{2}}}\\
F^{3\prime} &= \frac{F^{3}\sqrt{1-\frac{v^{2}}{c^{2}}}}{1-\frac{u_{1}v}{c^{2}}}
\end{aligned}$$

If, in particular, the particle happens to be instatntaneously at rest in the frame S, the force transformations become particularly simple

$$\begin{aligned}
F^{1\prime} &= F^{1}\\
F^{2\prime} &= F^{2}\sqrt{1-\frac{v^{2}}{c^{2}}}\\
F^{3\prime} &= F^{3}\sqrt{1-\frac{v^{2}}{c^{2}}}
\end{aligned}$$

We will have occasion to use these transformations in a big way in the next chapter.

6.12 Using energy-momentum conservation

6.12.1 Inelastic collision, again

I will start with a generalized version of the perfectly inelastic collision that we discussed in section 6.5. This time the two particles that collide initially have masses (no longer identical) m_1 and m_2 and Doppler factors of k_1 and k_2 respectively. They stick together to produce a composite particle of mass μ and Doppler factor κ. Equations (6.21) and (6.22) lead to

$$\begin{aligned} \mu\kappa &= m_1 k_1 + m_2 k_2 \\ \frac{\mu}{\kappa} &= \frac{m_1}{k_1} + \frac{m_2}{k_2} \end{aligned}$$

which can be easily solved to find

$$\begin{aligned} \mu &= \sqrt{(m_1 k_1 + m_2 k_2)\left(\frac{m_1}{k_1} + \frac{m_2}{k_2}\right)} \\ \kappa &= \sqrt{\frac{m_1 k_1 + m_2 k_2}{\frac{m_1}{k_1} + \frac{m_2}{k_2}}} \end{aligned}$$

Check that in the special case $m_1 = m_2 = m$ and $k_1 = k_2^{-1} = k$, this does reduce to the the values $\mu = 2m\gamma$ and $\kappa = 1$ as expected.

6.12.2 Decay and stability

Next, I will show you the consequences of energy and momentum conservation for a very real process - the α decay of a heavy nucleus

$$_Z X^A \rightarrow \alpha +_{Z-2} Y^{A-4}$$

If we write down the equation for the mass-energy balance for this equation, what we get is

$$E(X) = E(\alpha) + E(Y)$$

or, in terms of the kinetic and mass energies,

$$m_X c^2 = m_\alpha c^2 + T(\alpha) + m_Y c^2 + T(Y)$$

where I have obviously chosen to write the equation in the frame in which the parent nucleus is it rest before the decay (So that $T(X) = 0$). Thus the net kinetic energy of the decay products is given by

$$T(\alpha) + T(Y) = (m_X - m_Y - m_\alpha) c^2 = Q \tag{6.48}$$

and is thus equal to the mass excess that the decaying nucleus has over the decay products, apart from a factor of c^2. Note that the factor of c^2 arises only if we insist on using conventional units for measuring mass - in nuclear physics, mass is usually measured in energy units like the MeV[7], in which case, the mass difference is directly equal to the final kinetic energy. It should be easy for you to see that if the frame being used is *not* the frame in which X is at rest, the the mass difference gives the increase in net kinetic energy in the process, $T(Y) + T(\alpha) - T(X)$. As I have written in (6.48), the mass difference is denoted by Q. Why Q? The analogy with heat released by a chemical reaction should make this clear!

What I have said just now remains equally valid for all decay

[7]The Mega-electronVolt, or 10^6 eV. Remember, an eV is the amount of kinetic energy gained by an electron accelerated by a potential difference of 1 Volt, and is equal roughly to 1.6×10^{-19} Jin terms of conventional units.

processes. For example, the free neutron decays by the process

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e} + \overline{\nu}$$

for which, the final kinetic energies in the rest frame of the neutron obeys[8]

$$T(\mathrm{p}) + T(\mathrm{e}) + T(\overline{\nu}) = \left(m_{\mathrm{n}} - m_{\mathrm{p}} - m_{\mathrm{e}} - m_{\nu}\right) c^2$$

Similarly, in a nuclear beta decay,

$$_{Z}X^{A} \rightarrow_{Z+1} Y^{A} + \mathrm{e} + \overline{v}$$

the net kinetic energy of the decay products is given by

$$T(Y) + T(\mathrm{e}) + T(\overline{\nu}) = \left(m_X - m_Y - m_{\mathrm{e}} - m_{\overline{\nu}}\right) c^2$$

In general, in any decay process, where a mother particle X decays into a set of particles $P_1, P_2, \ldots, P_n$ you can easily see that we can write

$$T(P_1) + \ldots T(P_n) - T(X) = \left(m_X - m_{P_1} - \ldots - m_{P_n}\right) c^2$$

where the Q value of this reaction is given by $\left(m_X - m_{P_1} - \ldots - m_{P_n}\right) c^2$ - the mass difference times c^2. If we work in the frame where the mother particle X is at rest, then this is of course the net kinetic energy of all the decay particles.

Let us now think of a process very much like the decay of the free neutron - a decay of a free proton into a neutron, a positron

[8] I have included the mass of the neutrino in the expression, simply because current research shows that it is probably non-zero. However, the upper bound that we have on the neutrino mass is so very small, compared to even the mass of a light particle such as the electron that it can be safely ignored in these expressions.

and a neutrino[9]

$$p \rightarrow n + e^{+} + \nu$$

The Q value of this decay process is negative as is easily seen from the fact that the neutron alone is more massive than the proton (indeed, since the neutron decay process has a positive Q value, it is more massive than the proton, the electron and the neutrino put together!). In the rest frame of the proton, this implies that the net kinetic energy of the decay products is negative - which is absurd! The conclusion, then, is that, at least in the rest frame of the proton, this decay can not occur!

It may strike you that this problem can be overcome by simply speeding up the proton so that its kinetic energy is sufficient to overcome the deficit of the negative Q value. Can this be done? All you have to realize is that you can just as well analyze the situation from the point of view of someone who is moving at the same speed as the proton. In this frame all laws of physics are equally valid, and here the proton is at rest and so can not decay according to the argument given above. Now, if the proton does not decay in one frame it does not decay in any other frame either! Thus, speeding up the proton can not make it decay.

You may feel a bit cheated by the argument against proton decay presented above - you could feel it more honorable to be able to conclude the same while staying in the frame in which the proton is moving very fast. However, the beauty of the whole argument lies in the fact that we can judiciously choose the frame to simplify it. Indeed this is one feature that makes applying STR a joy - the

[9]The reason why you need to have a positron here in this case, instead of the electron, is because the net electrical charge has to balance out on the two sides. The third particle is the neutrino, rather than the anti-neutrino because of a conservation law called the lepton number conservation - but all this will take us too far afield.

principle of relativity allows you to use any frame whatsoever, while a particular frame may make the calculation (or the argument) very simple.

"All that is very well," you could still say, "but couldn't we have come to the same conclusion sticking to the old frame?" The answer to this is - we could, but only aftera slightly more complicated argument. The point is, not only does the photon's decay have to conserve energy - it also has to conserve momentum. Since the proton ahs a huge momentum in our frame, the net momentum of the three decay products has to be huge too. Thus, the final kinetic energies not only have to be positive, they have to be very large in order to ensure that they have the requisite momentum. As we can show, the extra kinetic energy needed for this is actually greater than the kinetic energy that the proton had in the first place - so the demand of conservation of momentum more than offsets the advantage that the high initial speed could have brought in!

Thus, it is energy momentum conservation that ensures that the proton can not decay in this particular manner. However, the natural question is - why can not the proton ecay into something less massive - there are certainly many, many particles that can play the role of possible decay products? The answer lies in the realms of particle physics - but without going too far afield I can tell you that it has to do with a law called baryon number conservation. Roughly, it says that a baryon (like protons, neutrons, etc.) can only decay into a baryon. Now, the proton happens to be the lightest among the baryons - making it impossible for the proton to decay[10].

[10]Impossible, perhaps, is too strong a word to use here. If baryon number conservation is an exact law of physics then, of course, there is no possibility of proton decay. However, some theories of particle physics (the so called grand unified theories) allow for the possibility of a very small violation of this law -

We can make the above into a general law - a particle cannot decay into products that are, together, more massive than it. On the other hand, if a proposed decay scheme obeys all conservation laws and has a positive Q value, to boot - then the decay will occur! Note though that the rate at which the decay will occur can not be completely predicted from STR (quantum mechanics plays a huge role) it is true that a larger Q value does enhance the decay rate.

6.12.3 Dividing up the energy

We can go a lot further than just predict the total kinetic energy of the decay products. After all, there is one more conservation law, the conservation of momentum (or rather, as we have seen, one more portion of the law of conservation of the momentum four vector) that I have not made use of as yet! For the case of the α decay, momentum conservation allows us to write

$$p\left(Y\right) + p\left(\alpha\right) = 0 \tag{6.49}$$

Now, the typical kinetic energies in an α decay process is a few MeV - which is much less than the rest energies of the particles involved. This means that we can safely use non-relativistic expressions for the kinetic energy and momentum. Hence (6.49) becomes

$$2m_{\alpha}T\left(\alpha\right) = 2m_{Y}T\left(Y\right)$$

opening up a tiny possibility of proton decay. However, experiments running over the last decade have not been able to reveal any sign of proton decay. Having said that, one can not rule out the possibility that it occurs, but with a such a tiny probability that is below the threshhold of our current detection schemes.

which means that (6.48) can be written

$$T(\alpha)\left(1+\frac{m_\alpha}{m_Y}\right)=Q$$

leading to a precise prediction for the αparticle kinetic energy

$$T(\alpha)=\frac{Q}{1+\frac{m_\alpha}{m_Y}} \tag{6.50}$$

To get an idea about the sort of energies that we are talking about here, let me consider a real example - the α decay of U^{238} into Ra^{234}.

Note that the derivation of (6.50) was made very simple by the fact that the energies involved were small enough to ensure that non-relativistic approximations could be used. What if the speeds involved are much larger (as happens, for example, in beta decay - with electrons emerging at nearly the speed of light)?

A direct calculation of the energies carried away by the two decay products can be carried out along the lines of the non-relativistic calculations above. However, the algebra does tend to become a little messy. Below we describe two ways of getting the result - the first by using our k-calculus equations and the second by using the energy-momentum 4-vector.

6.12.3.1 The k-calculus approach

Let the Doppler factors of the two particles in the rest frame of the parent particle be k_1 and k_2, respectively. In this case, the equations (6.21-6.22) become

$$M=m_1k_1+m_2k_2=\frac{m_1}{k_1}+\frac{m_2}{k_2}. \tag{6.51}$$

Eliminating k_2 between these two equations leads to the quadratic equation

$$k_1^2 + \left(\frac{m_2^2 - m_1^2 - M^2}{m_1 M}\right) k_1 + 1 = 0 \tag{6.52}$$

for k_1. Of course, we can solve this this for the Doppler factor (and hence the speed) of the first particle. However, if all we want the energy $E_1 = \frac{m_1 c^2}{2}\left(k_1 + \frac{1}{k_1}\right)$, all we have to do is merely rearrange (6.52) to get

$$k_1 + \frac{1}{k_1} = \frac{M^2 + m_1^2 - m_2^2}{m_1 M} \tag{6.53}$$

and thus

$$E_1 = \frac{m_1 c^2}{2} \times \frac{1}{m_1 M}\left(M^2 + m_1^2 - m_2^2\right) = \frac{c^2}{2M}\left(M^2 + m_1^2 - m_2^2\right) \tag{6.54}$$

Of course, the interchange $m_1 \leftrightarrow m_2$ gives E_2 as

$$E_2 = \frac{c^2}{2M}\left(M^2 - m_1^2 + m_2^2\right) \tag{6.55}$$

One easily checks that this is consistent with the conservation of energy, $E_1 + E_2 = Mc^2$. The kinetic energies are given by

$$\begin{aligned} T_1 &= E_1 - m_1 c^2 = \frac{M - m_1 + m_2}{2M} Q && (6.56) \\ T_2 &= \frac{M + m_1 - m_2}{2M} Q && (6.57) \end{aligned}$$

where $Q = (M - m_1 - m_2)\, c^2$ is the total available kinetic energy.

The momentum is given by

$$|p_1| = \frac{m_1 c}{2}\left|k_1 - \frac{1}{k_1}\right|$$

and elementary algebra, once again, tells us from (6.53) that

$$|p_1| = \frac{m_1 c}{2} \times \sqrt{\left(\frac{M^2 + m_1^2 - m_2^2}{m_1 M}\right)^2 - 4}$$

which can be rearranged into the rather nice symmetric form

$$|p_1| = \frac{c}{2M}\sqrt{(M + m_1 + m_2)(M + m_1 - m_2)(M - m_1 + m_2)(M - m_1 - m_2)} \tag{6.58}$$

The symmetry shows that $|p_2|$ is the same, as it must be.

6.12.3.2 The 4-vector approach

There is a simple trick that uses the property of four-vectors to solve the problem. Let me now show you how it works. In this case, the law of four momentum conservation

$$P_X = P_Y + P_\alpha$$

can be rewritten as

$$P_Y = P_X - P_\alpha.$$

Taking the dot product of each side with itself leads to

$$P_Y^2 = P_X^2 + P_\alpha^2 - 2P_X \cdot P_\alpha$$

Note that this helps us get rid of the two unknown quantities involving the Y nucleus - its momentum and energy in one go! Now, $P_X^2 = m_X^2 c^2$, etc. Also,

$$P_X \cdot P_\alpha = \left(m_X c, \vec{0}\right) \cdot \left(\frac{E_\alpha}{c}, \vec{p}_\alpha\right) = m_X E_\alpha$$

which tells us that

$$m_Y^2 c^2 = m_X^2 c^2 + m_\alpha^2 c^2 - 2m_X E_\alpha$$

yielding the expression

$$E_\alpha = \frac{m_X^2 + m_\alpha^2 - m_Y^2}{2m_X}\, c^2$$

for the total, energy of the α particle. Of course, all you have to do to find the total energy of the Y nucleus after the decay is interchange the values of m_α and m_Y - which yields

$$E_Y = \frac{m_X^2 + m_Y^2 - m_\alpha^2}{2m_X}\, c^2$$

As you can easily verify, the sum of the energies of the decay products *is* equal to the rest energy of the original particle,

$$E_\alpha + E_Y = m_X c^2$$

To get the kinetic energies of the decay products, all you have to do is subtract the respective rest energies, yielding

$$T_\alpha = \frac{(m_X - m_\alpha)^2 - m_Y^2}{2m_X}\, c^2 \qquad (6.59\text{-a})$$

$$T_Y = \frac{(m_X - m_Y)^2 - m_\alpha^2}{2m_X}\, c^2 \qquad (6.59\text{-b})$$

You should be able to verify that these expressions reduce to the non-relativistic ones we found out earlier in the limit of low speeds.

Note that in this process the kinetic energy available for the decay products (which is nothing but the Q value) gets distributed among the two decay products in a well defined manner. We thus expect the α particle to come out with a definite energy - and that is precisely what has been observed[11]! Of course, this can be traced back to the fact that α decay is a two body process - the available kinetic energy gets shared between the daughter nucleus and the α particle in a precise ratio dictated by the conservation of momentum. The β decay spectrum, on the other hand, is another story altogether - instead of a precise energy (or a set of sharp lines) the β particles come out with all energies from zero upto a maximum. Indeed, the maximum kinetic energy of the β particles (called the **endpoint energy** of the β spectrum) turns out to be what (6.59-a) would predict (Of course with m_α replaced by m_β). Of course, the explanation is very simple - β is not a two body process - the available Q value is shared between the three particles in the final state! Thus, the β particle can come out with any energy below the endpoint value - the rest is carried away by the neutrino, while maintaining conservation of momentum. Note that the fact that a two body decay could not explain the continuous decay spectrum in beta decay is the precise reason why Wolfganf Pauli had proposed the existence of a third decay product - the neutrino! Indeed, it

[11]Actually, if you look at the α decay spectrum, which is nothing but a plot of the number of α particles emitted against their respective energies, what you seeussally is not one, but several sharp lines. Thus, the αparticles come out with not one definite energy, but several well defined energies. Explaining this is not too difficult, thoough - it is just that the daughter nucleus is not always produced in its ground state. Thus the energy term that I wrote down foer the daughter nucleus is not just $m_Y c^2 + T_Y$, but has an extra term of $\mathcal{E}_Y^*$ as well, where $\mathcal{E}_Y^*$ is the excitation energy of the state the nucleus is created in. As you can easily check, this brings down the available kinetic energy from Q to $Q - \mathcal{E}_Y^*$ - which explains where the α particles of energies lower than that predicted by (6.50) come from. In fact, these α particles of lower eneries are accompanied with γ rays (photons) that carry away the extra energy that becomes available when the excited daughter nucleus de-excites to the ground state - this provides us with evidence that this explanation is correct! Since quantum mechanics tells us that the excited states of the daughter nucleus have a few definite energies only, there are only a few sharp lines in the α spectrum.

took thirty years for the neutrino to be experimentally detected after Pauli's prediction.

Today, the idea that we need an extra particle to ensure that energy-momentum conservation works out correctly may not seem to be too far-fetched. After all, there are hundreds of so called elementary particles, so what's the big deal about one particle more? You should realize, though, that when Pauli first proposed the neutrino to explain the continuous spectrum in beta decay, the number of known elementary partices could be counted on the fingers of your hand - the particle explosion was yet to take place! Adding one more particle to the list just to ensure that energy-momentum conservation holds good is a prime example of "changing the universe" to suit theory - the closest analogy that I can think of is the proposal of the existence of Neptune to explain deviations in the orbit of Uranus from the calculated path. While the latter emphasises the degee of confidence that eighteenth century physicists had in Newton's laws, the fact that very few doubted the existence of the neutrino in the thirty intervening years between the neutrino's prediction and discovery is testimony to our degree of faith in energy-momentum conservation!

6.12.4 Nuclear reactions

Energy-momentum conservation has a major role to play in nuclear physics. I have already shown you the major role it plays in nuclear decays, this time I will talk about its role in deciding when a nuclear reaction will occur.

How is a nuclear reaction different from a decay? A decay has only one particle in the initial state. In a reaction, several particles (I am using the word particle in an extended sense, heavy nuclei count as particles, too!) take part in the initial state - but the most common situation is where two particles react - *e.g.* when you shoot α particles at a nucleus in order to cause a transmutation. In most cases, you shoot a projectile particle (which may be an α particle, a proton ...) at a stationary target containing the nuclei you want to transmute. Of course, I am describing this from a particular frame of reference - the frame in which the target is at rest. this is typically called the lab frame.

The concept of the "heat of reaction" Q holds good for reactions too. In this case, its value is c^2 times the net mass of the reactants minus the net mass of the products of the reaction. Conservation of energy tells us that this will be the increase in the net kinetic energy of the products over those entering into the reaction. In this case, we borrow terminology from chemistry and call a reaction exothermic if $Q > 0$, and endothermic if $Q < 0$. Note that a decay will never occur if $Q < 0$ for it, but it is possible to have a reaction which is endothermic - all that will happen is that the products will emerge with less kinetic energy than that carried in by the reactants. Of course, the incoming particles must have enough energy to make-up the deficit.

If the Q value of an endothermic reaction

$$X + \alpha \rightarrow Y_1 + Y_2 + \ldots$$

is, say, -4.0 MeV, you may be inclined to believe that the the reaction will occur if you shoot α particles with a minimum energy of 4.0 MeV at a stationary target of X nuclei. In reality, the α particles have to be more energitic than this. The reason is the same as the one that prevented the fast moving proton from decaying - in this case, the fact that the incoming α particle has some momentum ensures that the decay products would also have to have some minimum kinetic energy to begin with - just providing them with enough to overcome the Q deficit will not do!

Just providing enough kinetic energy to overcome the Q deficit will be enough to cause the reaction - but only if you can carry it under the most favourable situation for the reaction to occur - one in which the products can be produced at rest! Conservation of momentum tells you that this is possible only if the initial momentum is zero (this is why we looked at the rest frame of the mother particle in the case of a decay). Thus, the lab frame (in which the target has zero momentum but the projectile has a lot of it) is not suitable for looking at the extreme case - we should look at the so called center of momentum (CM) frame. For an endothermic reaction with $Q = -4.0$ MeV to occur, the net kinetic energy of the reactants in the CM frame must be at least +4.0 MeV. What is important for you to know, if you want to design an experiment to carry out this reaction, is the minimum kinetic energy that the projectile must have in the lab frame. Transforming from the CM frame to the lab frame and *vice versa* is thus very important in all studies of nuclear reactions. However in whhat follows, we will use two approaches, based on (what else!) the K-calculus and 4-momenta, respectively, to calculate the threshold energy directly in the lab frame.

6.12.4.1 The k-calculus approach

Let the Doppler factor of the projectile particle m_1 be k in the lab frame. In this case, the equalities (6.21) and (6.22) take the form

$$\begin{aligned} m_1 k + m_2 &= \sum_{i=1}^{N} \mu_j \kappa_j \\ \frac{m_1}{k} + m_2 &= \sum_{i=1}^{N} \frac{\mu_j}{\kappa_j} \end{aligned}$$

so that simply multiplying both sides gives us

$$\begin{aligned} m_1^2 + m_2^2 + m_1 m_2 \left(k + \frac{1}{k}\right) &= \left(\sum_{i=1}^{N} \mu_j \kappa_j\right)\left(\sum_{i=1}^{N} \frac{\mu_j}{\kappa_j}\right) \\ &= \sum_{j=1}^{N} \mu_j^2 + \sum_{j<k=1}^{N} \mu_j \mu_k \left(\frac{\kappa_j}{\kappa_k} + \frac{\kappa_k}{\kappa_j}\right) \end{aligned}$$

Now, $m_1 m_2 \left(k + \frac{1}{k}\right) = 2m_2 E_1/c^2$, while elementary algebra tells us that

$$\left(\frac{\kappa_j}{\kappa_k} + \frac{\kappa_k}{\kappa_j}\right) = 2 + \left(\sqrt{\frac{\kappa_j}{\kappa_k}} - \sqrt{\frac{\kappa_k}{\kappa_j}}\right)^2 \geq 2$$

with equality holding only for $\kappa_j = \kappa_k$. Thus

$$m_1^2 + m_2^2 + 2m_2 E_1 \geq \sum_{j=1}^{N} \mu_j^2 + \sum_{j<k=1}^{N} 2\mu_j \mu_k = \left(\sum_{j=1}^{N} \mu_j\right)^2 = M^2$$

where M is the net mass of the products of the reaction. The equality is reached, of course, when all the decay particles have the same k factor - *i.e.*, they move with as a single lump after the reaction. This inequality, which we will rederive using 4-momenta arguments below, allows us to put a lower bound on the energy of the incident projectile.

6.12.4.2 The 4-momentum approach

Another way to look at this hinges around the result for P^2 that is expressed in (6.30) and the inequality (6.31). If the net four-momentum of the products of the reaction is P, then we can write

$$P_1 + P_2 = P$$

where P_1 and P_2 are the four-momenta of the target and the projectile, respectively. Now, the two incident 4-momenta take the form $P_1 = (E_1/c, \vec{p}_1)$ and $P_2 = \left(m_2 c, \vec{0}\right)$ in the lab frame, where E_1 is the energy of the projectile in this frame - so that we have $P_1 \cdot P_2 = m_2 E_1$. "Squaring" both sides and of $P_1 + P_2 = P$, we get

$$c^2 \left(m_1^2 + m_2^2\right) + m_2 E_1 = P^2 \geq c^2 M^2$$

which is the same inequality that we have derived using the K-calculus above.

Since the kinetic energy of the projectile in the lab frame K_{lab} is given by $E_1 - m_1 c^2$, we must have

$$2m_1 K_{\text{lab}} \geq c^2 \left(M^2 - (m_1 + m_2)^2\right) = -Q\,(M + m_1 + m_2)$$

and thus and endothermic reaction becomes energitically feasible only if

$$K_{\text{lab}} \geq \frac{M + m_1 + m_2}{2m_1}\ |Q| = \left(1 + \frac{m_2}{m_1} + \frac{|Q|}{2m_1}\right)\ |Q| \approx \left(1 + \frac{m_2}{m_1}\right)\ |Q|$$

This shows that the lighter the target is, the harder the reaction gets. In fact, if an endothermic reaction is to occur by a projectile particle hitting an identical, stationary, particle, the projectile must have more kinetic energy than twice the Q deficit!

6.12.5 The Compton effect

We have already met the Comton effect in its most extreme form in section 6.3. There we talked about a photon bouncing back off a stationary electron. In general, though, the photon would bounce off at some angle from the incident direction. Since the photon gives the electron a kick, giving up some of its energy - it ends up with less energy than it started out with. This means that the photon emerging out of the collision will have a lower frequency and hence longer wavelength than when it came in. This is exactly what was observed by Arthur Compton in 1922. The Compton effect was the first direct observation of the particle nature of the photon so far as it proved that not only do photons carry the energy of light, they also carry momentum!

6.12.5.1 The shift

To calculate the shift in the wavelength of the photon, we start by writing down the conservation laws. The photon-electron collision is shown in figure .

We can immediately write down the law of conservation of energy

$$\hbar c k_i + mc^2 = \hbar c k_f + \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} \tag{6.60-a}$$

as well as the X and Y components of the law of conservation of energy

$$\hbar k_i = \hbar k_f \cos\vartheta + \frac{mv}{\sqrt{1-\frac{v^2}{c^2}}}\cos\phi \tag{6.60-b}$$

$$0 = \hbar k_f \sin\vartheta - \frac{mv}{\sqrt{1-\frac{v^2}{c^2}}}\cos\sin\phi \tag{6.60-c}$$

Since our aim is to find out the shift in the photon wavelength against its deflection angle ϑ, we must get rid of the electron speed

v and the angle ϕ. Eliminating ϕ from (6.60-b-6.60-c) gives

$$\frac{m^2v^2}{1-\frac{v^2}{c^2}} = \hbar^2\left(k_i^2 + k_f^2 - 2k_ik_f\cos\vartheta\right)$$

Again, (6.60-a) gives

$$\frac{m^2c^2}{1-\frac{v^2}{c^2}} = m^2c^2 + 2mc\hbar\left(k_i - k_f\right) + \hbar^2\left(k_i - k_f\right)^2$$

Subtracting these two equations immediately gives rise to

$$m^2c^2 = m^2c^2 + 2mc\hbar\left(k_i - k_f\right) - 2\hbar^2k_ik_f\left(1-\cos\vartheta\right)$$

leading to

$$\frac{\hbar}{mc}\left(1-\cos\vartheta\right) = \frac{1}{k_f} - \frac{1}{k_i} = \frac{1}{2\pi}\left(\lambda_f - \lambda_i\right)$$

This gives the formula

$$\Delta\lambda = \frac{h}{mc}\left(1-\cos\vartheta\right) \tag{6.61}$$

for the shift in wavelength in the Compton effect.

This formula fits the experimental observations completely - except for one small fact. Compton had found that along with a shift in wavelength as predicted above, the light scattered in a particular direction also had a component that had the same wavelength as the incident light. The question is, how could the photon deflect after kicking an electron without losing any of its energy? The answer lies in the m in the denominator of the equation above. A photon may collide with an electron that is so tightly bound with its atom that it inot the electron, but the entire atom that recoils in the collision. This means that the mass that should be used in the equation is that of the entire atom rather than that of the electron

- no wonder the shift in wavelength is too small to detect in this case.

6.12.5.2 The four-vector approach

Let me now show you how simple this calculation becomes if we make use of the properties of the energy-momentum four-vector. Note that in what we did above, my aim was to eliminate the speed and direction of motion of the electron in the final state. This can be immediately achieved by taking the norm of the energy-momentum four-vector for the electron - this will leave just m^2c^2, wiping out all traces of $\vec{v}$.

The conservation of four-momentum

$$P_{\gamma_i} + P_{e_i} = P_{\gamma_f} + P_{e_f}$$

can be rewriiten as

$$P_{e_f} = P_{\gamma_i} - P_{\gamma_f} + P_{e_i}$$

which when "squared" leads to

$$m^2c^2 = m^2c^2 + 2\left(P_{\gamma_i} - P_{\gamma_f}\right) \cdot P_{e_i} - 2\,P_{\gamma_i} \cdot P_{\gamma_f}$$

where I have used $P_{\gamma_i}^2 = P_{\gamma_f}^2 = 0$ and $P_{e_i}^2 = P_{e_f}^2 = m^2c^2$. This means that

$$P_{\gamma_i} \cdot P_{\gamma_f} = \left(P_{\gamma_i} - P_{\gamma_f}\right) \cdot P_{e_i}$$

Now, in the lab frame, we have

$$P_{\gamma_i} = \hbar\left(k_i, \vec{k}_i\right), \quad P_{\gamma_f} = \hbar\left(k_f, \vec{k}_f\right), \quad P_{e_i} = \left(mc, \vec{0}\right)$$

which means that the above equation becomes

$$\hbar^2\left(k_ik_f - \vec{k}_i\cdot\vec{k}_f\right) = \hbar\left(k_i - k_f\right)mc$$

From this, you can immediately arrive at the Compton shift formula, equation (6.61).

6.12.6 Relativistic billiards

As every billiards player knows, if a ball strikes another of the same mass at rest, after the collision, which is practically elastic, the two move off in mutually perpendicular directions. This result is valid for classical physics - here we wish to examine how it gets modified when the projectile ball is really fast - so fast that relativistic mechanics needs to be applied. A difference between this problem and the previous ones is that unlike in those cases no simple, elegant 4 -vector argument presents itself. The method which we will use here falls back directly on the principle of relativity itself: since the laws of physics are the same in all frames, solve the problem in a frame where the problem becomes simple and transform your results back to the frame in which you want the answer. To illustrate this method, let me use it to solve this problem in the non-relativistic case first[12].

[12] I do not claim that this is the simplest way to solve the non-relativistic version of this problem. Arguably the simplest solution could run as follows: conservation of momentum and kinetic energy will lead to the equations $\vec{v}_1+\vec{v}_2=\vec{u}$ and $v_1^2+v_2^2=u^2$ after cancellation of factors involving m. If you are geometrically minded you could simply say that the fisrt equation implies that the vector $\vec{u}$ forms the third side of a triangle formed by $\vec{v}_1$ and $\vec{v}_2$, while the second equation says that this triangle is right-angled! The more algebraically inclined might "square" the first equation to arrive at $u^2=\left(\vec{v}_1+\vec{v}_2\right)^2=v_1^2+v_2^2+2\vec{v}_1\cdot\vec{v}_2$ which implies $\vec{v}_1\cdot\vec{v}_2=0$!

6.12.6.1 The non-relativistic problem

Since this is a purely mechanical problem, we do have a principle of relativity to fall back upon in this case even for non-relativistic physics. The frame in which the problem simplifies considerably is the CM (center of momentum) frame where the net momentum vanishes. This means that in this frames the two balls approach each other with equal and opposite velocities $\vec{u}_{\text{CM}}$ and $-\vec{u}_{\text{CM}}$, respectively. After the collision, the net momentum should again be zero - which implies that the particles will again have equal and opposite velocities, $\vec{v}_{\text{CM}}$ and $-\vec{v}_{\text{CM}}$. Conservation of energy then tells us that $|\vec{u}_{\text{CM}}| = |\vec{v}_{\text{CM}}|$.

To find out the situation in the lab-frame all we have to do is apply the relevant Galilean transformation to the velocities. Since the second ball is initially at rest in the lab frame, all you have to do here is subtract the velocity $-\vec{u}_{\text{CM}}$ from every velocity. Thus, the post-collision velocities of balls 1 and 2 are $\vec{v}_1 = \vec{v}_{\text{CM}} + \vec{u}_{\text{CM}}$ and $\vec{v}_2 = -\vec{v}_{\text{CM}} + \vec{u}_{\text{CM}}$. This means that

$$\vec{v}_1 \cdot \vec{v}_2 = -\vec{v}^2_{\text{CM}} + \vec{u}^2_{\text{CM}} = 0$$

since $|\vec{u}_{\text{CM}}| = |\vec{v}_{\text{CM}}|$!

6.12.6.2 The relativistic case

the first part os the solution is identical for relativistic billiards. In the CM frame, the two balls do emerge with equal and opposite velocities with the same magnitude v as the initial velocities. The only reason why the relativistic case is somewhat more complicated is that the velocity transformations back to the lab frame are more complicated in the relativistic case.

Since all we are interested in are the directions of the post-collision velocities, the problem simplifies somewhat. In the CM frame, we take the initial direction of motion of the two particles as the X' axis. The two emerging particles make the angles θ' and $\phi' = \pi - \theta'$ to the X' axis in this frame. To find the corresponding angles, θ and ϕ in the lab frame, we can use the particle aberration formula to get:

$$\begin{aligned}\tan\theta &= \frac{v\sin\theta'}{\gamma(v)(v\cos\theta' + v)} = \frac{\sin\theta'}{\gamma(v)(\cos\theta' + 1)} \\ \tan\phi &= \frac{\sin\phi'}{\gamma(v)(\cos\phi' + 1)} = \frac{\sin\theta'}{\gamma(v)(-\cos\theta' + 1)}\end{aligned}$$

The simplest way to eliminate the unknown (and unwanted) angles θ' and ϕ' is to multiply these two equations to get

$$\tan\theta\,\tan\phi = \frac{1}{\gamma^2(v)}$$

Note that in the non-relativistic limit $v \ll c$, $\gamma(v) \to 1$, which tells us that $\tan\theta\,\tan\phi = 1$, and hence we recover the result $\theta + \phi = \pi/2$. In the relativistic case, however, the product $\tan\theta\,\tan\phi$ is less than one - implying that the angles between the final velocity directions is less than $\pi/2$.

We are not done with the relativistic case yet, though. We still have an unknown $\gamma(v)$ to get rid of, in favor of the known quantity $\gamma(u)$, where u is the speed of the projectile in the lab frame. Though we can calculate $\gamma(v)$ in terms of $\gamma(u)$ both directly or via the equation (5.24) we will use a 4-vector argument to directly derive this. The 4-velocities U_1 and U_2 have components of $\gamma(v)\left(c, \vec{v}_{\mathrm{CM}}\right)$ and $\gamma(v)\left(c, -\vec{v}_{\mathrm{CM}}\right)$, respectively, in the CM frame, while in the lab frame, the components are $\gamma(u)(c, \vec{u})$ and $\left(c, \vec{0}\right)$, respectively. Equating the values of the invariant $(U_1 + U_2)^2$ calculated in the two frames gives us

$$\begin{aligned} 4\gamma^2(v)\,c^2 &= (\gamma(u)+1)^2 c^2 - \gamma^2(u)\,u^2 \\ &= c^2\left(\gamma^2(u)\left[1-\frac{u^2}{c^2}\right]+2\gamma(u)+1\right) = 2c^2(\gamma(u)+1) \end{aligned}$$

so that

$$\gamma^2(v) = \frac{\gamma(u)+1}{2}$$

and our formula becomes

$$\tan\theta\,\tan\phi = \frac{2}{\gamma(u)+1} \tag{6.62}$$

This shows that the faster the projectile, the smaller $\tan\theta\,\tan\phi$ is, and hence the smaller the angle between the emergent directions. This is one of the predictions of relativity that can be directly tested in cloud chamber type experiments. Observing the tracks when a fast cosmic-ray electron hits a stationary electron shows this reduction in angle vividly. Indeed, by observing the angular distribution of the outgoing tracks a lot of information can be gathered about the incident electron's speed.

6.12.7 The relativistic rocket

Let us consider a rocket which propels itself by ejecting some of its mass backwards - thereby gaining speed by the conservation of momentum. Although the principle of the rocket is simplicity itself, the actual calculation is complicated, even for non-relativistic physics, by the fact that the rocket has a variable mass. Here we will calculate how the speed of a relativistic rocket varies with its mass. We will make the simplifying assumption that mass is ejected from the rocket at a constant speed $\mathcal{U}$ in the backward direction with respect to the rocket.We will once again solve the problem using two different approaches - one based on the K-calculus, the other a more conventional one.

6.12.7.1 The K-calculus approach

To simplify the notation, we make use of our old friends

$$\gamma(k) = \frac{1}{2}\left(k + \frac{1}{k}\right) \qquad \pi(k) = \frac{1}{2}\left(k - \frac{1}{k}\right)$$

from which it immediately follows that

$$\begin{aligned} \gamma(k) + \pi(k) &= k \\ \gamma(k) - \pi(k) &= k^{-1} \end{aligned}$$

We will write the constant K-factor of the ejected mass with respect to the rocket as $\mathcal{K}^{-1}$(since the mass moves backwards, its K-factor is less than 1 - or $\mathcal{K}$, however, is larger than 1), where

$$\mathcal{K} = \sqrt{\frac{c+\mathcal{U}}{c-\mathcal{U}}}$$

Now let the mass of the rocket change from M to $M + dM$ in the infinitesimal time interval from t to $t + dt$, in which time the rockets K-factor with respect to some fixed inertial frame changes from k to $k + dk$. In this frame, which we will take to be the frame in which the rocket is initially at rest for convenience, the K-factor of the ejected mass is $k/\mathcal{K}$. Let the mass of the ejected gas in this interval be δM (in classical physics, conservation of mass would have told us that $\delta M = -dM$. In relativity, however, it is energy that is conserved, and not mass!). Then, the laws of conservation of energy and momentum applied to this immediately leads to

$$\begin{aligned} M\gamma(k) &= (M+dM)\,\gamma(k+dk)+\delta M\,\gamma(k/\mathcal{K}) \\ M\pi(k) &= (M+dM)\,\pi(k+dk)+\delta M\,\pi(k/\mathcal{K}) \end{aligned}$$

which can be rewritten in the form

$$\begin{aligned} -\delta M\,\gamma(k/\mathcal{K}) &= dM\,\gamma(k)+M\,\dot\gamma(k)\,dk \\ -\delta M\,\pi(k/\mathcal{K}) &= dM\,\pi(k)+M\,\dot\pi(k)\,dk \end{aligned}$$

We can easily eliminate the unknown δM by dividing the two equations above to get

$$\frac{\gamma(k/\mathcal{K})}{\pi(k/\mathcal{K})}=\frac{k^2+\mathcal{K}^2}{k^2-\mathcal{K}^2}=\frac{dM\,\gamma(k)+M\,\dot\gamma(k)\,dk}{dM\,\pi(k)+M\,\dot\pi(k)\,dk}$$

and using componendo and dividendo on this yields

$$\frac{k^2}{\mathcal{K}^2}=\frac{dM\,\left(\gamma(k)+\pi(k)\right)+M\,\left(\dot\gamma(k)+\dot\pi(k)\right)dk}{dM\,\left(\gamma(k)-\pi(k)\right)+M\,\left(\dot\gamma(k)-\dot\pi(k)\right)dk}=\frac{k\,dM+M\,dk}{k^{-1}dM-k^{-2}M\,dk}$$

and thus

$$\frac{1}{\mathcal{K}^2}=\frac{k\,dM+M\,dk}{k\,dM-M\,dk}$$

so that

$$\frac{k}{M}\,\frac{dM}{dk}=\frac{1+\mathcal{K}^2}{1-\mathcal{K}^2}=-\frac{c}{\mathcal{U}}$$

which yields the differential equation

$$\frac{dM}{M}=-\frac{c}{\mathcal{U}}\,\frac{dk}{k}$$

Using the initial values $M=M_0$ when $k=1$, we can immediately solve this equation to get

$$\frac{M}{M_0} = k^{-c/\mathcal{U}} = \left(\frac{c+u}{c-u}\right)^{-c/2\mathcal{U}} \tag{6.63}$$

which tells us how the final mass decreases with increasing speed of the rocket. Thus, the higher the speed that you want your rocket to achieve, the smaller can the payload carried by your rocket be! To achieve a larger residual mass for a given final speed u, $\mathcal{U}$ must be as large as possible. This gives us the photon ship - very popular in science fiction - which propels itself by ejecting a stream of photons! For a photon ship (6.63) becomes

$$\frac{M}{M_0} = \sqrt{\frac{c-u}{c+u}} \tag{6.64}$$

6.12.7.2 The conventional approach

Instead of using the K-calculus we can use the same physical arguments directly in terms of the velocity. The velocity of the ejected mass δM in the initial rest frame of the rocket is given by $\overline{\mathcal{U}} = (u - \mathcal{U}) / (1 - u\mathcal{U}/c^2)$. This is where the reason behind the simplicity of the K-calculus approach should become evident! We can bravely press forward and write the energy and momentum conservation equations as

$$\begin{aligned} Mc^2\gamma(u) &= (M + dM)\, c^2\gamma(u + du) + \delta M c^2 \gamma\left(\overline{\mathcal{U}}\right) \\ M\gamma(u)\, u &= (M + dM)\, \gamma(u + du)(u + du) + \delta M \gamma\left(\overline{\mathcal{U}}\right) \overline{\mathcal{U}} \end{aligned}$$

from which it follows that

$$\overline{\mathcal{U}} = \frac{d\,(M\gamma(u)\, u)}{d\,(M\gamma(u))} = u + M\gamma(u)\,\frac{du}{d\,(M\gamma(u))}$$

This implies

$$\frac{1}{M\gamma(u)}\frac{d}{du}(M\gamma(u)) = \frac{1}{\overline{\mathcal{U}} - u} = -\frac{1 - u\mathcal{U}/c^2}{(1-u^2/c^2)\mathcal{U}} = -\frac{1}{2\mathcal{U}}\left(\frac{1-\mathcal{U}/c}{1-u./c} + \frac{1+\mathcal{U}/c}{1+u/c}\right)$$

leading to the solution (using $M = M_0$ at $u = 0$)

$$\ln\frac{M\gamma(u)}{M_0\gamma(0)} = \ln\frac{M}{M_0} + \ln\gamma(u) = -\frac{c}{2\mathcal{U}}\left[\ln\frac{1+u/c}{1-u/c}\right] - \frac{1}{2}\ln\left(1-u^2/c^2\right)$$

which simplifies to

$$\frac{M}{M_0} = \left(\frac{c-u}{c+u}\right)^{c/2\mathcal{U}}$$

which is the solution we were after.

There is an arguably slightly simpler way of arriving at this solution. This involves using a special frame in which the description of the process becomes simple. This frame is the ICF, the frame in which the rocket is stationary at a particular instant of time. At a time t, the ICF will move at a velocity u with respect to the initial rest frame of the rocket. After a time dt, the ejected gas (mass δM) will have a velocity $-\mathcal{U}$ while the rocket will have a velocity of du'(Remember, the ICF is *not* fixed to the rocket - and thus the rocket is stationary in this frame only at the time t). Thus, energy and momentum conservation in this frame leads to

$$\begin{aligned} Mc^2 &= (M+dM)\gamma(du')c^2 + \delta M\gamma(-\mathcal{U})c^2 \\ 0 &= (M+dM)\gamma(du')du' + \delta M\gamma(-\mathcal{U})(-\mathcal{U}) \end{aligned}$$

Noting that $\gamma(du') = \gamma(0) + \dot{\gamma}(0)du' + \mathcal{O}(du'^2) = 1 + \mathcal{O}(du'^2)$ we get $\delta M\,\gamma(-\mathcal{U}) = -dM$ and thus

$$M du' = -dM\,\mathcal{U} \tag{6.65}$$

Unfortunately, we cannot directly integrate du' to get the increase in the rockets velocity. This is because the increments du' at different times are with respect to different frames. In order to find the velocity, we need to find out velocity increments for a single frame - say, the initial rest frame of the rocket. We can use the velocity addition law

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

to find

$$du = \frac{(1 + u'v/c^2)\, du' - (u' + v)\, v du'/c^2}{(1 + u'v/c^2)^2} = \frac{(1 - v^2/c^2)\, du'}{(1 + u'v/c^2)}$$

Since in the ICF, $v = u$ and $u' = 0$, this means that

$$du = \left(1 - \frac{u^2}{c^2}\right) du' = -\left(1 - \frac{u^2}{c^2}\right) \mathcal{U} \frac{dM}{M}$$

which can be easily solved to get (6.63). If you ask me, I find the simplicity that one gains by going over to the ICF is ruined by the fact that you haveto convert du' back to du before you can integrate it. Make your own choice!

As far as the photon rocket is concerned, we have already derived the result (6.64) for it as a special case of the general result (6.63). There is, however, a much simpler way of deriving this result, and this hinges on the the fact that the energy - momentum relation for photons is linear, which means that the net momentum carried away by the photons is just the total energy of the photons divided by c (since for a massive particle we have $|\vec{p}| = \sqrt{E^2 - m^2c^4}/c$, the total momentum of the gas, ejected at different speeds with respect to a given inertial observer at different times, can not be directly related to the total energy in a simple way for a rocket ejecting massive particles. In other words, you can not calculate the total momentum from the total energy of a bunch of massive particles - while you can do this for a gas of photons!). This allows us to directly use energy-momentum conservation between the initial and the final stages of the rocket to get (E_γ is the net energy of the photon gas)

$$\begin{aligned} E_\gamma + M\gamma(u)\,c^2 &= M_0c^2 \\ p_\gamma + M\gamma(u)\,u &= 0 \end{aligned}$$

Since $E_\gamma = -p_\gamma c$ (remember, p_γ is negative and thus $|\vec{p}| = -p_\gamma$), the two equations combine to give

$$M\gamma(u)\left(1+\frac{u}{c}\right) = M_0$$

which immediately leads to (6.64).

6.13 Motion under force

So far we have made do with just the conservation laws of energy and momentum. This is in stark contrast with the classicalapproach to elementary mechanics - where force and Newton's second law plays a central role.

Chapter 7

Electrodynamics and relativity

7.1 Faraday and Einstein

7.2 Why does a current produce a magnetic field?

Imagine that the history of physics in some other world is quiet different from the one in our's. In this world, the special theory of relativity has come in a lot earlier - before the fact that currents produce magnetic fields has been discovered! All that is known of what we call electromagnetism is Coulomb's law for electrostatics. In this imaginary world, let us follow the footsteps of a pioneering physicist who is trying to deduce the effect of a currrent in a straight wire on a charge moving in its vicinity.

7.3 Transforming the fields

7.4 The field of moving charges

7.5 Potentials to the fore

7.6 Light - again!