Introduction to Special Relativity

E. Hardy

edward.hardy@physics.ox.ac.uk

Rudolf Peierls Centre for Theoretical Physics, University of Oxford,
Parks Road, Oxford, OX1 3PU, UK

The parts of the text that are coloured red are strictly non-examinable. They give, for those who are interested, a more advanced perspective. The black text is not necessarily all examinable (the syllabus is definitive), but it is material that I think you ought to know to understand the subject properly and be well-prepared for your subsequent years and, eventually, research.

Contents

1 Introduction **2**

2 Coordinate systems and inertial reference frames **5**

3 Galilean relativity **6**

4 Electromagnetism **13**

5 Principles of special relativity **18**

6 The end for universal time **19**

7 The invariant interval and Lorentz transformations **20**

8 Digression: natural units **27**

9 Causality **29**

10 Spacetime diagrams **34**

11 Time dilation and length contraction **35**

12 Proper time **41**

13 Addition of velocities **44**

14 "Paradoxes" **47**

15 4-vectors **48**

16 4-velocity and 4-acceleration **52**

17 Relativistic energy and momentum **57**

18 Massless particles and Doppler shift **62**

19 Centre of momentum frame **65**

20 Relativistic collisions and decays **67**

21 Newton's laws **79**

22 Conclusions and what next? **83**

A Time dilation and length contraction again **84**

B Examples of "paradoxes" **84**

1 Introduction

Newtonian mechanics works well in many everyday situations. However, it gives predictions that are experimentally wrong when applied to systems that are moving too fast (with speeds v close to the speed of light c, so $v/c \ll 1$ is violated) or are too small (e.g. of size ℓ comparable to the Compton wavelength $\hbar/(mc)$ of a particle of mass m, so $\ell \gg \hbar/(mc)$ is violated) or for which the gravitational field is strong (such that spacetime is not close to flat, with $GM/R \ll c^2$ violated). In these cases we must turn to other theories, respectively special relativity, quantum mechanics, and general relativity.[1] All of these are more fundamental than Newtonian mechanics, which is at best only a good approximation (note that, e.g., even if $v \ll c$ the corrections to Newtonian mechanics from special relativity can be important if you want a very precise prediction). Interestingly, each of these theories forces us to change our view of how the Universe works quite dramatically: with special relativity we lose the idea of universal time, quantum mechanics is intrinsically probabilistic, and general relativity tells us that spacetime is curved and leads to black holes.

A natural question is what happens when several of the conditions for Newtonian mechanics to be valid are simultaneously violated. General relativity subsumes special relativity so covers the case of large gravitational fields and high speeds. Combining quantum mechanics and special relativity leads to quantum field theory (taught in a 4th year course), which retains the key principles of both as well as leading to remarkable new predictions (anti-particles, the connection between spin and whether particles are bosons or fermions, etc.). We do not know for sure how to combine general relativity and quantum field theory; string theory is arguably the most promising possibility but as yet this is not theoretically complete or experimentally tested.

Useful exercise: Calculate the mass of a particle for which the gravitational potential at a distance of its Compton wavelength is of order c^2. Find the Compton wavelength of such a particle, in units of metres. This is known as the Planck length and sets the typical length scale below which quantum gravity is expected to be relevant. You could discuss the implications for the testability of string theory with your tutors.

In this course we give an introduction to special relativity. Special relativity is arguably the first "modern" physics theory that you are being taught. This is true in the simple sense that it's much more recent than your other courses so far, although it is still more than a century old. Much more importantly, the approach of thinking about symmetries of theories that led to special relativity still remains at the forefront of theoretical physics. Moreover, special relativity is, to the best of our knowledge, basically correct. By this I mean that the key principles of special relativity survive pretty much unchanged in our best current theories, to the extent that if a candidate theory of elementary particle physics or cosmology does not respect Lorentz invariance (described later), it can usually immediately be discarded. Of course, it might be that one day an

[1] General relativity is actually also required if gravitational effects are relevant and $v/c \sim 1$ even if the spacetime curvature is small.

experiment discovers a small deviation from special relativity that necessitates we reconsider or a compelling theory might be found that predicts a tiny deviation, either of which would cause a revolution but would require extraordinary evidence.[2]

The aim of this course is two-fold. One purpose is to get you familiar with the basics of special relativity and to demonstrate some of the interesting and experimentally important implications. The vast majority of this can be done using only fairly simple maths. A second, perhaps less obvious, purpose is to introduce you to some of the mathematical structures and ways of thinking that, although not essential here, are indispensable when you reach more advanced courses.

I use the convention that the invariant interval is $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$ rather than $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$. This is common in special relativity and particle physics, but in general relativity you will often come across the opposite choice (for example, your third year courses will use $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$). I retain factors of c, see Section 8 below. Three dimensional vectors are denoted in bold, e.g. $\mathbf{x}$, and 4-vectors are generally capitalized, e.g. X.

References and literature

There are many books dedicated to special relativity, and it is worth looking at a few to find one with a style that you like. For example, you might want to try

- Special Relativity, by A.P. French (CRC Press)
- Introducing Special Relativity, by W.S.C. Williams (Taylor and Francis)
- Special Relativity, by N.M.J. Woodhouse (Springer)
- Relativity Made Relatively Easy, by A.M. Steane (OUP)

The discussions of special relativity in books focused on other, more advanced, topics can also be useful. I especially recommend

- Classical Electrodynamics, by J.D. Jackson (Wiley): Chapters 11 and 12, which can be read independently of the preceding chapters
- The Classical Theory of Fields, by L.D. Landau and E.M. Lifshitz (Pergamon Press): early chapters
- A First Course in General Relativity, by B. Schutz (Cambridge University Press): early chapters

[2] See `https://thereader.mitpress.mit.edu/when-science-fails-opera-neutrinos/` for a recent incident.

Lecture notes for other courses can be useful

- The MIT course 8.033 on relativity is at the right level but has more time than us so covers some additional interesting topics and gives an introduction to general relativity: `https://ocw.mit.edu/courses/8-033-relativity-fall-2006/pages/lecture-notes/`
- The lecture notes by Prof. Christopher Beem for a course in the Oxford Mathematics Department are at about the correct level but with a slightly different style: `https://courses.maths.ox.ac.uk/course/view.php?id=4951`
- As always, the lecture notes by Prof. David Tong for a Cambridge course are useful: `https://www.damtp.cam.ac.uk/user/tong/relativity/seven.pdf`
- The notes and especially the many questions for a course by Prof. Masud Haque at the Max Planck Institute might be useful: `https://www.pks.mpg.de/~haque/maynooth_MP352/`

I thank Prof. Caroline Terquem and the previous lecturers of this course, most recently Prof. Kimberley Palladino.

2 Coordinate systems and inertial reference frames

We start with some definitions that set the stage on which the physics will take place.

A ***reference frame*** is a system of coordinates covering all of space (you can picture a grid with tick marks that enable positions to be measured) along with clocks fixed to this grid from which time can be measured. We take the definition of a reference frame to include a choice of spatial origin $x = y = z = 0$ (i.e. a tick mark that is the origin at all times) and a choice of a zero time $t = 0$ synchronised across all the clocks. I sometimes use the notation S or similar to denote a specific reference frame.

An ***event*** is a particular point in space and time (i.e. it has infinitesimal spatial extent and duration). ***Spacetime*** is the set of all events. Relative to a given reference frame S, we can assign spatial and time coordinates to any specific event: (t, x, y, z) say. We package the time and space coordinates into a single vector and follow the convention of writing the time coordinate first. The numerical values of the coordinates assigned to a particular event differ between different reference frames, e.g. in another frame S' the same event might have coordinates (t', x', y', z'). One of our key tasks in this course is to determine how the coordinates in different reference frames of some event are related, i.e. the ***transformation*** between reference frames.

A postulate of both Newtonian mechanics and special relativity is the existence of reference frames in which a freely moving body, i.e. one that is not acted on by any external forces, travels with constant velocity. These are known as ***inertial reference frames***.[3] I will use the term ***inertial coordinate system*** interchangeably with inertial reference frame (sometimes people take only the latter to have a zero time defined).

Another postulate of both Newtonian physics and special relativity is that if two reference frames are moving relative to each other with constant velocity and one is inertial then so is the other. This is plausible: a uniformly moving particle in one such frame might be expected to move uniformly also in the other. In this way, given one inertial frame we can construct infinitely many others. A frame that is accelerating relative to an inertial frame is not inertial.

Let me finish this section with a small caution. When we talk about the coordinates of an event in a particular reference frame, we are imagining a team of infinitely many helpful assistants all in the same reference frame, one at each point in space, ready to record when things happen. In contrast, if we want to know when a specific person in a particular frame *sees* an event happening, we also need to account for the time it takes for light from the event to reach them (which is non-zero in special relativity). This is an extra layer of complication compared to just determining the coordinates of the event, and for most of our purposes is not what we are interested in (e.g. we assume experimentalists have removed this effect from their measurements already).

[3] We assume that inertial reference frames exist and can be defined globally, i.e. across all of spacetime. This is a postulate of both Newtonian physics and special relativity, but is not true in general relativity, for which typically only local inertial frames exist.

3 Galilean relativity

Before special relativity, a principle of relativity had already been formulated, originally by Galileo. He wrote about this in terms of "fish, butterflies, and other small flying animals" on ships, but let me modernise the thought experiment slightly.

> *Imagine a spaceship travelling through empty space. If the ship is accelerating, then passengers can feel this. However, in the absence of acceleration, there is no physical test that could be done to decide whether it is at rest or moving. In other words, no device within the spaceship can detect uniform motion, only motion relative to some external standard (e.g. the Earth).*[4]

As a result, given two spaceships moving at uniform relative velocity it is impossible for the two sets of passengers to decide which are moving and which are stationary. Both are equally entitled to claim that they are the ones at rest. The principle of relativity turns this thought experiment from a negative to a positive statement.

Principle of relativity

The laws of physics hold in any inertial frame; in particular they take the same *form* when written in terms of quantities measured in any inertial frame.

Consequently, it only makes sense to talk about a reference frame being in motion *relative* to another one.

Subsequently, we will almost exclusively consider inertial reference frames, so if I write "frame" without qualification I mean an inertial frame.

The principle of relativity will remain with us in special relativity and its consequences have been experimentally verified to high precision. It is powerful because it restricts the form of physical laws, dramatically narrowing down the range of theories that we must consider as candidates to describe the Universe.

Usually ***Galilean relativity*** is understood as the application of the principle of relativity to Newton's laws (which, as discussed, are only approximations that are accurate if all speeds are small). In this context, we also assume that time progresses at the same rate in all reference frames (i.e. time is universal). This means there is a "universal clock" with at most the definition

[4] In the real Universe, the Cosmic Microwave Background (leftover photons from the hot Big Bang) provides a particularly natural external standard. This does not conflict with the principle that in totally empty space no experiment can distinguish being rest from uniform motion. In general, it's important to separate the properties of a particular solution of the equations of a system, of which our Universe with the Cosmic Microwave Background is one example, from the properties of the equations themselves.

of $t = 0$ shifted between different frames. We also assume that the mass of any object is the same in all frames.

How are the coordinates (t, x, y, z) of an event measured in an inertial frame S related to the coordinates (t', x', y', z') of the same event measured in another frame S' in Galilean relativity? We could actually derive the transformation rule from the requirement that Newton's second law takes the same form in both frames and that time is universal. However, since the result matches up to our intuition and it will be superseded by special relativity anyway, let me just state it: If S' is moving with velocity $\mathbf{v} = (v_x, v_y, v_z)$ as viewed in S and the two frames synchronise their origins and their clocks at $t = 0$, and the alignment of their axes, then

$$\begin{aligned} t' &= t \ , \\ x' &= x - v_x t \ , \\ y' &= y - v_y t \ , \\ z' &= z - v_z t \ . \end{aligned} \tag{1}$$

To make sure the signs in Eq. (1) are correct, consider the coordinates of the spatial origin of S in S'. The transformation Eq. (1) is called a ***Galilean boost***.

If the relative velocity is in the direction of the x-axis $\mathbf{v} = (v, 0, 0)$, Eq. (1) can be written in the compact form

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \ , \qquad y' = y, \qquad z' = z \ . \tag{2}$$

The inverse transformation, from S' to S, is obtained simply by replacing $\mathbf{v}$ with $-\mathbf{v}$. We can check this for a boost in the x-direction by calculating its effect on the S' coordinates of an event

$$\begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} \begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} t \\ x \end{pmatrix} \ , \tag{3}$$

as required (it is also clear from Eq. (1)).

One can additionally easily check that first applying a transformation from S to S', with relative velocity $\mathbf{v}_1$, then from S' to a third inertial frame S'' with relative velocity $\mathbf{v}_2$ is equivalent to going directly from S to S'' with a transformation of velocity $\mathbf{v}_1 + \mathbf{v}_2$. Finally, there is an identity transformation corresponding to $\mathbf{v} = 0$.

Inspired by these properties, let us take a more abstract view of transformations between frames. Consider a set of elements $G = \{g_i\}$ one for each possible boost, together with a composition law $g_i \cdot g_j = g_k \in G$ specifying the effect of two successive transformations (which must itself be a transformation). In the case of Galilean boosts this composition law is simple: in obvious notation $g_{\mathbf{v}} \cdot g_{\mathbf{w}} = g_{\mathbf{v}+\mathbf{w}}$. There is also an identity element and, for each transformation, an inverse. Together these properties make $\{g_i\}$ a *group*. Once we consider rotations as well as boosts we

obtain another, more interesting, group. In this case the composition law is more complicated, e.g. the order that we apply a rotation and a boost matters so $g_i \cdot g_j \neq g_j \cdot g_i$ in general.

Example: Galilean transformations
A person standing on the Earth throws a ball up in the air with speed w and catches it on its way back down. In the (approximately) inertial reference frame in which the Earth is stationary, and with coordinate axes aligned such that the z-direction is pointing vertically upwards and a suitable time and space origin, the trajectory of the ball is $z(t) = wt - \frac{1}{2}gt^2$ for $t \in [0, 2w/g]$. What is the trajectory of the ball seen in a reference frame S' moving vertically upwards relative to the Earth's rest frame with speed v? At what time is the ball at rest in this reference frame?

Assume $t = 0$ and $t' = 0$ are synchronised. From Eq. (2), $z'(t') = wt' - \frac{1}{2}gt'^2 - vt'$ with $t' \in [0, 2w/g]$. The ball is at rest in this frame when $dz'(t')/dt' = 0 \implies t' = (w - v)/g$.

Two inertial reference frames can also be related by a spatial rotation. In this case

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R \begin{pmatrix} x \\ y \\ z \end{pmatrix} \ , \tag{4}$$

where R is an orthogonal matrix $R^T R = I$ (with determinant $+1$ since we want to keep the handedness of the coordinate system unchanged). Two inertial reference frames can also be related by a redefinition of the origin of time or space, i.e. a translation

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} + \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} \ , \tag{5}$$

where c_i are constants. The most general transformation between two inertial frames is the combination of a Galilean boost, a rotation and a translation in space and time (shifts of the spatial and temporal origins).[5]

Let us now return to check that Newton's laws actually do take the same form in any two inertial frames. We start with an example.

Example: Newtonian Gravity and Galilean boosts
Consider two particles of mass m_1 and m_2 interacting gravitationally. In a frame S' the first

[5] There are actually also transformations that rescale our coordinate system e.g. $\mathbf{x}' = \lambda \mathbf{x}$ for $\lambda \in \mathbb{R}$ and analogously for t. These aren't interesting for our purposes, but turn out to be extremely important in some areas of physics (e.g. systems near a phase transition are approximately invariant under such transformations).

particle's position $\mathbf{r}'_1 = (x'_1, y'_1, z'_1)$ evolves according to

$$m_1 \frac{d^2\mathbf{r}'_1}{dt'^2} = -\frac{\mathbf{r}'_1 - \mathbf{r}'_2}{|\mathbf{r}'_1 - \mathbf{r}'_2|} \frac{Gm_1m_2}{|\mathbf{r}'_1 - \mathbf{r}'_2|^2} \ , \tag{6}$$

(the first factor on the right-hand side just gives a unit vector in the direction of the force). Consider now a frame S, such that S' is moving relative to S with speed $\mathbf{v}$. Using Eq. (1), the left-hand side of Eq. (6) can be written in terms of $\mathbf{r}_1$

$$\frac{d^2\mathbf{r}'_1}{dt'^2} = \frac{d^2\mathbf{r}'_1}{dt^2} = \frac{d^2(\mathbf{r}_1 - \mathbf{v}t)}{dt^2} = \frac{d^2\mathbf{r}_1}{dt^2} \ . \tag{7}$$

Also,

$$\mathbf{r}'_1 - \mathbf{r}'_2 = (\mathbf{r}_1 - \mathbf{v}t) - (\mathbf{r}_2 - \mathbf{v}t) = \mathbf{r}_1 - \mathbf{r}_2 \ . \tag{8}$$

Hence, rewriting Eq. (6) in terms of the coordinates of the frame S

$$m_1 \frac{d^2\mathbf{r}_1}{dt^2} = -\frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|} \frac{Gm_1m_2}{|\mathbf{r}_1 - \mathbf{r}_2|^2} \ , \tag{9}$$

which is of exactly the same form as in S'. Therefore, Newton's law of gravity is obeyed in both S and S'.

We can easily extend the calculation in the example above to any force arising from a potential $V(\mathbf{r}'_1 - \mathbf{r}'_2)$ that depends only on the distance between the two particles, not their absolute position. For the x-component of the position of particle 1, we have

$$m_1 \frac{d^2x'_1}{dt'^2} = -\frac{d}{dx'_1} V(\mathbf{r}'_1 - \mathbf{r}'_2) \ , \tag{10}$$

and the equations for the other components of the position and for the second particle are similar (once you meet the operator ∇ next term you'll be able to write these equations more concisely). We immediately have $V(\mathbf{r}'_1 - \mathbf{r}'_2) = V(\mathbf{r}_1 - \mathbf{r}_2)$, and

$$\frac{d}{dx'} = \frac{dx}{dx'}\frac{d}{dx} = \frac{d}{dx} \ , \tag{11}$$

so

$$\frac{d}{dx'_1} V(\mathbf{r}'_1 - \mathbf{r}'_2) = \frac{d}{dx_1} V(\mathbf{r}_1 - \mathbf{r}_2) \ , \tag{12}$$

and so on. Substituting this and Eq. (7) into Eq. (10), the equation of motion takes the same form in terms of the S or S' coordinates.

From these calculations we see that I was a bit too fast before! The principle of relativity puts additional constraints on the allowed form of potentials. For example, a potential that depends on the position of a particle rather than the difference between two positions would not satisfy the principle of relativity (in practice, this would arise from a particle interacting with some

background e.g. a ball rolling up a hill, such that the principle of relativity is restored once we consider the background to also be part of the system).

Under a Galilean boost, not only does the *form* of Newton's second law stay the same but the *numerical values* of the acceleration and force are the same in two frames, see Eq. (7). This second fact is not a requirement of the principle of relativity as the next example illustrates.

Example: Newtonian Gravity and rotations

Suppose we take our system of two gravitating particles evolving according to Eq. (6) in S', and transform to a frame S that is related to S' by a spatial rotation. The coordinates in the two frames are related by Eq. (4). Therefore Eq. (6) becomes:

$$\begin{aligned} m_1 \frac{d^2(R\mathbf{r}_1)}{dt^2} &= -\frac{R\mathbf{r}_1 - R\mathbf{r}_2}{|R\mathbf{r}_1 - R\mathbf{r}_2|} \frac{Gm_1m_2}{|R\mathbf{r}_1 - R\mathbf{r}_2|^2} \\ \implies m_1 R \frac{d^2\mathbf{r}_1}{dt^2} &= -\frac{R(\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \frac{Gm_1m_2}{|\mathbf{r}_1 - \mathbf{r}_2|^2} \\ \implies m_1 \frac{d^2\mathbf{r}_1}{dt^2} &= -\frac{(\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \frac{Gm_1m_2}{|\mathbf{r}_1 - \mathbf{r}_2|^2} \ , \end{aligned} \tag{13}$$

where in the second line we used that

$$|R\mathbf{r}_1 - R\mathbf{r}_2| = \sqrt{((R(\mathbf{r}_1 - \mathbf{r}_2))^T R(\mathbf{r}_1 - \mathbf{r}_2)} = \sqrt{(\mathbf{r}_1 - \mathbf{r}_2)^T R^T R(\mathbf{r}_1 - \mathbf{r}_2)} = |\mathbf{r}_1 - \mathbf{r}_2| \ , \tag{14}$$

(rotations preserve distances between points), where T is the transpose, and in the third line applied R^{-1} to both sides.

Not surprisingly, the acceleration and the force vectors are both different in the two frames ($d^2\mathbf{r}_1/dt^2 \neq d^2\mathbf{r}'_1/dt'^2$). This is fine, the key thing is that the *form* of the equation of motion is the same written in terms of either set of coordinates, so observers in both frames agree on the laws of physics. For this to be the case the two sides of the equation of motion had to (and did) transform in the same way under a rotation.

The previous examples motivate the following definitions:

Invariance and Covariance

- A physical quantity is Galilean ***covariant*** if it transforms in a definite way under Galilean transformations. For example, acceleration is covariant.
- There are some physical quantities that take the same numerical values in all inertial frames (these are covariant with a very simple transformation: they don't change). Such quantities are called ***scalars***, and are ***invariant*** under Galilean transformations.
- An equation is Galilean ***covariant*** if it can be written in terms of Galilean covariant quantities, and all terms transform in the same way.
- If an equation is Galilean covariant, then its *form* is Galilean invariant (to show this rewrite it in terms of the transformed variables then act on each term with the inverse transformation, as in Eq. (13)). Such equations therefore respect the principle of relativity: if they hold in one inertial frame, they hold in any inertial frame.

Caution: the word covariant is also used to describe a vector that transforms in a specific way, a "covariant vector" (as opposed to a "contravariant vector"). We won't need this in our present course but you'll meet it in your third year.

Let's be more precise about what a physical quantity transforming in a "definite" way means. In the previous red comment we described the transformations between frames as elements of a group. Suppose a physical quantity, (acceleration, event coordinate, etc.) can be represented as a vector $\mathbf{x}$. Then we require that each transformation g_i must act on $\mathbf{x}$ as some matrix: $D(g_i)\mathbf{x}$. We also require that the effect of acting with two transformations successively is consistent with the group composition $D(g_2)D(g_1) = D(g_2 \cdot g_1)$. Mathematically this means that the physical quantity transforms according to a (matrix) *representation* of the group of transformations.

The simplest example of a Galilean invariant is the mass of a particle. Another, which we have seen already, is the spatial distance between two events E_1 and E_2 that occur at the same time

$$t_1 = t_2 \implies (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 = (x_1' - x_2')^2 + (y_1' - y_2')^2 + (z_1' - z_2')^2 \ . \quad (15)$$

We define the length of an object to be the distance between two events at equal times, one at each end of the object. From Eq. (15) this is Galilean invariant. Another quantity that is obviously Galilean invariant is the time between two events.

Many physically interesting quantities are not invariant, e.g. the energy or momentum of a particle (acceleration is unchanged by Galilean boosts and translations but not rotations). Likewise, the distance between two events at different times is not invariant (so the statement that two non-simultaneous events occur at the same point in space has meaning only when we specify the reference frame used). *Short exercise: prove these claims.*

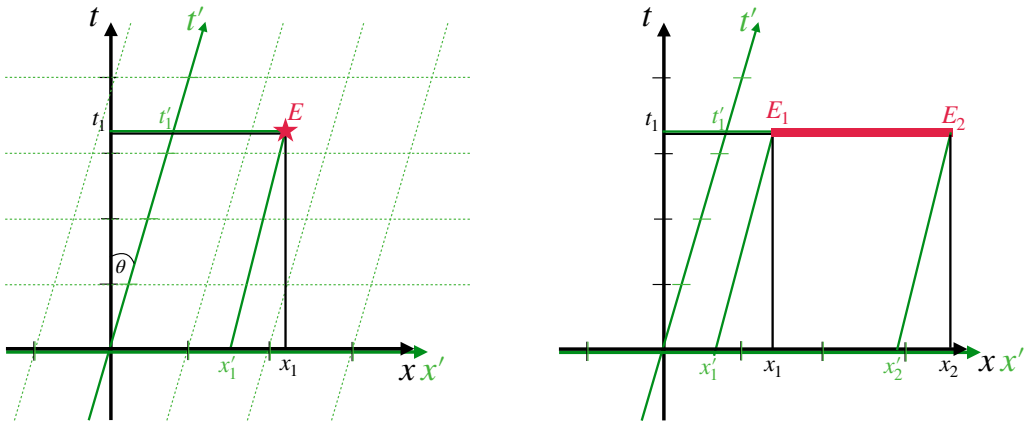


Figure 1: **Left:** A Galilean spacetime diagram for two frames S and S' related by a boost with speed v in the x-direction. The t' axis is inclined at angle $\theta = \arctan(\tilde{v})$ to the t-axis, where $\tilde{v}$ is the numerical value of the velocity in the units of the x and t axes. The coordinates of an event E in S are read off as for a normal graph, and the coordinates in S' are read by following the dotted lines. **Right:** The length of an object (red line) is defined as the spatial distance between two events E_1 and E_2 at either end of the object at the same time. In S the length is $\ell = x_2 - x_1$ and in S' it is $\ell' = x'_2 - x'_1$. We can easily check that $\ell = \ell'$.

Although overkill in the present context, in anticipation of what follows let us introduce a graphical approach to relating the coordinates of events in different frames. Consider two frames S and S' related by a Galilean boost in the x-direction as in Eq. (2). Ignoring the y and z coordinates, we can represent the coordinates of an event in S, (t, x), on a normal graph with orthogonal x, t axes.

On such a graph, the spatial origin of S', $x' = 0$, corresponds to the line $t = x/v$. This is the t'-axis. More generally, lines of constant x' are straight lines

$$t = \frac{x}{v} + \frac{x'}{v} \ . \tag{16}$$

Since time is universal, $t' = t$, and the x'-axis (i.e. the line $t' = 0$) coincides with the x-axis. Therefore, given some event E with coordinates (t, x) in S we can find the corresponding coordinate in S' as follows: we have $t' = t$ (of course) and to find the x'-coordinate we follow the coordinate lines to the x' axis. An example is shown in Fig. 1 left, and in the right panel we show how the length of an object is defined.

Note: in this course we think of transformations as moving between different reference frames with the physical system itself unchanged. These are known as ***passive transformations****. In other contexts it can be more useful to think of transformations as changing the physical system itself, not moving to a different reference frame. These are known as* ***active transformations****. Starting from one solution of the equations of motion an active transformation generates other*

solutions. Passive and active transformations are very closely related, with the main practical difference usually being minus signs, but it's important to be clear what perspective you're taking.

4 Electromagnetism

Special relativity was first discovered in the context of electromagnetism. This is perhaps not surprising given that light travels at the speed of light and is relatively straightforward to do experiments with (gravitational waves also travel at the speed of light but are much less experimentally convenient).

An immediate hint that trying to combine electromagnetism and the principle of relativity might lead to something interesting comes from the force law for a charged particle, in standard notation

$$\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \ . \tag{17}$$

The velocity $\mathbf{v}$ of the particle changes between reference frames, so for the principle of relativity to hold something non-trivial must happen. The situation looks tricky for Galilean relativity because the left-hand side is invariant under boosts, so the right-hand side must be too.

Actually, electromagnetism is not Galilean invariant. To see this note that in the absence of sources Maxwell's equations (which you will learn next term) have solutions with the x and z-components of $\mathbf{E}$ zero and the y-component E_y satisfying

$$\left(\frac{1}{c^2} \left. \frac{\partial^2}{\partial t^2} \right|_x - \left. \frac{\partial^2}{\partial x^2} \right|_t \right) E_y = 0 \ , \tag{18}$$

where c is the speed of light. The solution of Eq. (18) is

$$E_y = f(x - ct) + g(x + ct) \ , \tag{19}$$

where f, g are twice differentiable but otherwise arbitrary functions, corresponding to right and left moving waves of speed c. For example $E_y = \sin(k(x - ct)) = \sin(kx - \omega t)$, with $\omega = ck$, is a right moving sinusoidal wave. Such solutions, which involve $\mathbf{E}$ (and $\mathbf{B}$) orthogonal to the direction of propagation, are light waves travelling at speed c.

Suppose a ray of light travels in the positive x-direction in a frame S passing through events with coordinates satisfying $x = ct$. Consider a frame S' moving at speed v in the x-direction relative to S . Then if the Galilean boost is the correct way to relate the coordinates in the two frames, the light's trajectory in S' is

$$\begin{aligned} x' + vt' &= t'c \\ \implies x' &= (c - v)t' \ . \end{aligned} \tag{20}$$

Eq. (20) is just the "common sense" expectation that in S' the light travels at speed $v - c$. But this implies that Eq. (18) and therefore Maxwell's equations do not hold in S', which violates the principle of relativity.

Example: Galilean transformations and the wave equation

The fact that the form of Eq. (18) is not invariant under Galilean transformations is crucial. Therefore, let's show how transforming Eq. (18) to S' leads to the same conclusion. To do so, we use Eq. (6) and

$$\begin{aligned} \left.\frac{\partial}{\partial t}\right|_x &= \left.\frac{\partial t'}{\partial t}\right|_x \left.\frac{\partial}{\partial t'}\right|_{x'} + \left.\frac{\partial x'}{\partial t}\right|_x \left.\frac{\partial}{\partial x'}\right|_{t'} &&= \left.\frac{\partial}{\partial t'}\right|_{x'} - v \left.\frac{\partial}{\partial x'}\right|_{t'} \\ \left.\frac{\partial}{\partial x}\right|_t &= \left.\frac{\partial t'}{\partial x}\right|_t \left.\frac{\partial}{\partial t'}\right|_{x'} + \left.\frac{\partial x'}{\partial x}\right|_t \left.\frac{\partial}{\partial x'}\right|_{t'} &&= 0 + \left.\frac{\partial}{\partial x'}\right|_{t'} \ . \end{aligned} \tag{21}$$

We also make the reasonable assumption that the value of E_y at a particular spacetime event is unchanged by a boost in the x-direction, i.e. $E'_y(t', x') = E_y(t, x)$. Then Eq. (18) becomes

$$\left(\frac{1}{c^2}\left.\frac{\partial^2}{\partial t'^2}\right|_{x'} - \left(1 - \frac{v^2}{c^2}\right)\left.\frac{\partial^2}{\partial x'^2}\right|_{t'} - \frac{2v}{c^2}\left.\frac{\partial}{\partial x'}\right|_{t'}\left.\frac{\partial}{\partial t'}\right|_{x'}\right) E'_y = 0 \ . \tag{22}$$

Evidently the form of the equation has changed, so Maxwell's equations are not Galilean covariant, and do not satisfy the principle of Galilean relativity!

As a further cross check, in an optional question on the first problem sheet you will show that Eq. (22) has solutions of the form

$$\begin{aligned} E_y &= f(x' + vt' - ct') + g(x' + vt' + ct') \\ &= f(x' - (c - v)t') + g(x' + (c + v)t') \ . \end{aligned} \tag{23}$$

From the second line, we see that these solutions are waves that move to the right with speed $(c - v)$ and to the left with speed $(c + v)$. In contrast, if Maxwell's equations, and therefore also Eq. (18), held in all frames then light waves would have speed c in all frames. The solutions in Eq. (23) are just the Galilean transformation of Eq. (19), i.e. waves travelling at speed $c \pm v$, as expected from Eq. (20).

Eq. (20) is exactly how waves that are transmitted by a medium, such as sound waves, are seen to propagate by an observer moving relative to the medium. In these cases, the wave equations analogous to Eq. (18) (but for e.g. pressure rather than electric field) are only expected to hold in the rest frame of the medium. This is consistent with Galilean relativity because sound waves are understood as a consequence of Galilean classical mechanics and the existence of a preferred reference frame, that of the medium, is clear.

When physicists first noticed this situation with Maxwell's equations, there was no known

supporting medium for light. If this was the case then Maxwell's equations should instead hold in all inertial frames, which as we have just seen is inconsistent with Galilean relativity. Several possible resolutions of this problem were proposed:

1. Maxwell's equations are incorrect, and the true theory of electromagnetism is covariant under Galilean transformations.

2. There is an as yet undetected supporting medium for electromagnetic waves, the *luminiferous ether*. As a result, there is a preferred reference frame for electromagnetism, which is the frame in which the ether is at rest, and Maxwell's equations would hold only in this frame.

3. There exists a relativity principle for both classical mechanics and relativity but it was not Galilean relativity. In this case the laws of mechanics would need to be modified.

The first option was not appealing; remarkable agreement between experiments and the predictions of Maxwell theory seemed too good to ignore.

The second alternative was a bit worrying since the ether seemed truly ethereal, having no manifestation or effects other than supporting the propagation of light. Nevertheless, it was widely accepted by physicists at the time. Initial experimental efforts to observe the motion of the Earth relative to the ether, most famously the Michelson-Morley experiment (reviewed below) failed. However, various theoretical modifications were suggested. At one stage it was popular to assume that the ether was dragged along with the Earth. Another idea, by FitzGerald and Lorentz, was that objects moving at velocity $\mathbf{v}$ through the ether are contracted in the direction of motion by a factor of

$$\sqrt{1-\frac{|\mathbf{v}|^2}{c^2}}\,, \tag{24}$$

due to the impact of the ether on the electrons and atoms that comprise the object. This explains the null result of the Michelson–Morley experiment, and (as we will see) is the correct factor albeit from the wrong underlying physics. Eventually, with results from increasingly precise and ingenious experiments requiring more and more contrived theoretical patch-ups, the ether hypothesis became untenable.

In retrospect it's tempting to think of option 2 as obviously misguided. But, despite it ultimately being proved wrong by experiment, I think this is unfair. Several times throughout history physicists have invented new, previously unobserved, particles/ fields/ etc. to enable valued principles to be retained, and sometimes they were proved correct by later experiments. For example, Pauli invented the neutrino to keep energy/ momentum/ angular momentum conservation, and the Higgs boson was invented to retain key features of the Standard Model of particle physics. There have, of course, also been failures! At present in particle theory we are facing a somewhat similar challenge. We expect a principle known as "naturalness" to be obeyed, but the absence of beyond Standard Model physics discoveries at the Large Hadron Collider (LHC) at CERN is

putting this in tension. As yet it's unknown whether there is a sufficiently clever "patch-up", or whether we should discard this principle.

Example: The Michelson–Morley experiment

Suppose that the ether exists and is stationary with respect to the Sun. (These days we know the Sun is orbiting our galaxy, the galaxy is moving through the Local Group etc. but let's stick with the original picture.) Detecting or excluding the effect of the Earth's motion through such an ether is challenging because the fractional change in the speed of light is of order $v_{\rm E}/c$, where $v_{\rm E} \simeq 10^{-4}c$ is the speed of the Earth relative to the Sun. Nevertheless, Michelson and Morley developed a remarkable experiment that reached the required sensitivity.

A beam of light is split by a partially silvered plate. The two resulting beams travel perpendicularly, both for a distance of approximately L before being reflected. On reaching the plate again a component of each beam is directed to a detector. The recombination of the two beams leads to an interference pattern.

Let's call the lengths of the two arms L_1 and L_2. The interference pattern depends on the relative phase of the two beams when they return. If the Earth were at rest with respect to the ether then the travel time along the two arms would be $2L_1/c$ and $2L_2/c$. The phase difference between the returning light beams would therefore be $(2\pi)2(L_1 - L_2)/\lambda$, where $\lambda \sim 500\,\text{nm}$ is the wavelength of the light.

If an ether exists the time difference between the two paths, and therefore the relative phase, would be different due to the Earth's motion relative the ether. Suppose the experiment is carried out with arm 1 aligned with the motion of the Earth through the ether at speed $v_{\rm E}$. During the outward leg of the light's trip to the first mirror it is travelling against the ether wind, so is slowed down in the Earth's frame. Conversely, during the return trip it is sped up in the Earth's frame, so the round trip takes a total time

$$t_1 = \frac{L_1}{c - v_{\rm E}} + \frac{L_1}{c + v_{\rm E}} = \frac{2L_1}{c}\gamma(v_{\rm E})^2 \ . \tag{25}$$

where

$$\gamma(v) := \frac{1}{\sqrt{1 - v^2/c^2}} \ . \tag{26}$$

In case you haven't come across this notation, I use := to mean that the object on the left is defined by the expression on the right (vice versa for =:). Like most people, I'm not careful to always distinguish this from = and instead only use it where emphasis is valuable.

Meanwhile, the light going along arm 2 travels perpendicular to the ether. In Galilean relativity time is universal so we can evaluate the travel time in the ether rest frame (in which the light travels at speed c). By Pythagoras' theorem, the light travels a total distance

$$2\sqrt{L_2^2 + (v_{\rm E}t_2)^2} \ , \tag{27}$$

so the transit time t_2 satisfies

$$\begin{aligned} ct_2 &= 2\sqrt{L_2^2 + (v_{\rm E} t_2)^2} \\ \implies t_2 &= \frac{2L_2}{c}\gamma(v_{\rm E}) \ . \end{aligned} \tag{28}$$

The time difference between the two paths is therefore

$$\begin{aligned} \Delta t &\equiv t_2 - t_1 \\ &= \frac{2L_2}{c}\gamma(v_{\rm E}) - \frac{2L_1}{c}\gamma(v_{\rm E})^2 \ . \end{aligned} \tag{29}$$

There is no hope of knowing L_1 and L_2 to within an error of order λ, which is necessary to directly measure the part of the phase difference due to the Earth's motion relative to the ether, i.e. the part due to $\gamma(v_{\rm E}) \neq \gamma(v_{\rm E})^2$ in Eq. (29).

The crucial idea is that the experiment can be carefully rotated, so the role of the two arms is interchanged. Having done so, arm 2 is now aligned with the Earth's velocity through the ether. The time difference between the paths is then

$$\Delta\tilde{t} = \frac{2L_2}{c}\gamma(v_{\rm E})^2 - \frac{2L_1}{c}\gamma(v_{\rm E}) \ . \tag{30}$$

Consequently, the change in the difference of path times due to rotating the experiment is

$$\Delta\tilde{t} - \Delta t = 2\gamma(v_{\rm E})(\gamma(v_{\rm E}) - 1)\frac{L_1 + L_2}{c} \simeq \frac{L_1 + L_2}{c}\left(\frac{v_{\rm E}}{c}\right)^2 \ , \tag{31}$$

where we used that $\gamma(v) \simeq 1 + \frac{1}{2}v^2/c^2$ for $v \ll c$. We no longer need to know L_1 or L_2 precisely to detect the impact of the Earth's motion relative to the ether.

Eq. (31) means that when the experiment is rotated the phase ϕ of the interference pattern should change by

$$\delta\phi = (2\pi)\frac{c}{\lambda}\left(\Delta\tilde{t} - \Delta t\right) \simeq (2\pi)\frac{2L}{\lambda}\left(\frac{v_{\rm E}^2}{c^2}\right)^2 \ , \tag{32}$$

where λ is the frequency of the light. In the original experiment, $L \simeq 10\,\text{m}$ and $\lambda \simeq 500\,\text{nm}$ so the predicted $\delta\phi/(2\pi) \simeq 0.4$. Accounting for systematic uncertainties, the apparatus was sensitive enough to observe a shift $\delta\phi/(2\pi) \simeq 0.01$. This is a spectacular precision, corresponding to a change in the light travel time of order $10^{-17}\,\text{s}$.

The experiment gave a null result, with no phase shift detected when the experiment was rotated. Since then, various improved versions have been carried out, including e.g. using light from the Sun to remove the possibility that the motion of the source could affect the result.

Exercise: estimate the uncertainty in the measured phase shift in a modern day Michelson-Morley interferometer due to vibrations from a nearby road. What about systematic shifts due to, e.g., the temperature of the apparatus varying through the day?

5 Principles of special relativity

Einstein chose the third option above, and sought a relativistic principle that would encompass classical mechanics and electromagnetism. Experiment subsequently proved this to be correct! The result is special relativity, which is based on two postulates:

Einstein's postulates of relativity

1. The laws of physics take the same form in any inertial frame.
2. The speed of light is finite and, measured in any inertial reference frame, takes the same value in vacuum.

Postulate 1 is the same as the principle of relativity that we introduced in Galilean relativity. In practice, Postulate 2 means that Maxwell's equations take the same form for all inertial observers. Postulate 2 immediately explains the null result of the Michelson-Morley experiment: in the reference frame of the interferometer the time for light to travel along the two arms is independent of their orientation so rotating does not cause a phase shift.

Because special relativity applies to everything, not just light, it is desirable to replace the second postulate with an alternative that stresses the generality of the theory:

2'. In every inertial frame there is a finite, universal, limiting speed C for physical entities.

Experimentally the limiting speed C is found to be equal to the speed of light in a vacuum c (as we will see later, this is equivalent to the statement that photons are massless). Below we will derive the Lorentz transformations from postulates 1 and 2, but we could also derive them from 1 and 2', with the speed of light entering only to fix the numerical value of the limiting speed.

What would the implications be if it was discovered that the photon has a mass? There are in fact theories that allow this and are consistent with all present observations provided the mass is small enough ($\ll 10^{-33}\,\mathrm{eV}$). The formulation in terms of postulate 2' shows us that nothing dramatic would change, at least for special relativity. Like any massive particle, there would exist an inertial frame in which a given photon is at rest, so we could not identify C with the "speed of light". However, still no particle could be accelerated beyond the speed C (which would be extremely close numerically to what we currently call c).

Let us now claim an important fact, which will be justified later:

A maximum speed

In special relativity, the speed u of a particle with non-zero mass is always strictly less than the speed of light (i.e. $u < c$) in all inertial reference frames. A massless particle always travels at the speed of light in all inertial reference frames ($u = c$).

More generally, no *information* can be transmitted between observers at faster than the speed of light. If the sun suddenly exploded, we would have about 8 minutes (in our rest frame) until we knew about it.

It is important to distinguish information being transmitted from the results of measurements being *correlated*. Suppose someone sends two snooker balls both of the same colour to two distant observers. The moment that one looks at theirs, they know the colour of the other one. But no information has been transmitted in the moment of looking. Be aware however, you cannot think of this classical physics example as a model for quantum entanglement, which is more subtle although still the result of correlations.

In passing, let me note that we consider only systems in which gravity is not relevant (this is the case for, e.g., many particle physics experiments). As mentioned in footnote 1, if a system involves speeds not much smaller than c and gravitational effects are relevant then special relativity is not sufficient and instead general relativity is required.

6 The end for universal time

The postulates of special relativity require that we abandon the idea of universal time, as can be seen from the following thought experiment.

Consider two inertial reference frames S and S' with S' moving with speed v in the x-direction relative to S, as shown in Fig. 2. Suppose a light signal is emitted in all directions from a point A on the x-axis at some moment. Now consider two points in space, B and C, that are equidistant from A in opposite directions along the x-axis (and stationary in S). The emitted signal will reach B and C at the same time in S.

However, the two events corresponding to the arrival of the signal at B and C will not be simultaneous in the S' system. By the principle of relativity, the speed of the signal in the S' coordinate system has the same value c in both the forward and backward direction. Since the point B moves away from the source of the light signal, while the point C moves towards it, in the S' system we expect the signal to reach the point C before the point B.

From this, we learn the following:

The relativity of simultaneity in special relativity

In special relativity events that are simultaneous in one inertial reference frame S are not simultaneous in other inertial frames (specifically, in frames related to S by a boost). We say that simultaneity is relative, i.e. frame-dependent rather than absolute.

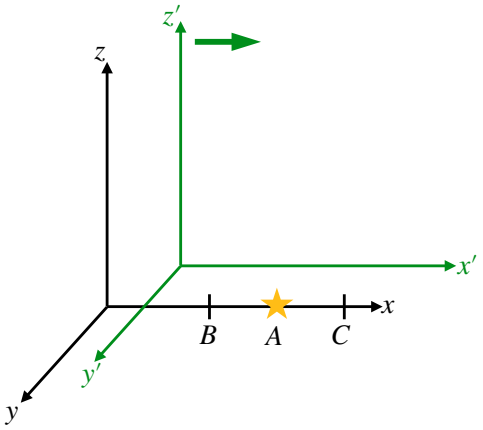


Figure 2: Setup of the thought experiment. A light signal is emitted from point A. In the frame S the signal arrives at the points B and C (which are stationary in S) simultaneously. In the frame S' the signal arrives at C before it arrives at B.

This is a major conceptual change compared to Newtonian mechanics. Our intuitive notions of space and time, based on everyday experience, are only approximations linked to the fact that we happen to deal only with speeds that are very small compared with the speed of light.[6]

7 The invariant interval and Lorentz transformations

We now derive a formula for the transformations between inertial frames in special relativity. We focus on ***Lorentz transformations***, which are transformations between two frames S and S' that have synchronised spacetime origins (i.e. the event with coordinate $(t, x, y, z) = (0, 0, 0, 0)$ in S has coordinate $(t', x', y', z') = (0, 0, 0, 0)$ in S'). There are various derivations. We follow the approach in Landau & Lifshitz (originally by Born I believe).

In addition to Einstein's postulates as stated above, we need one more postulate:

Homogeneity and isotropy

We assume that space and time are homogeneous (the same everywhere) and isotropic (the same in all directions).

Note that homogeneous does not imply isotropic: a universe containing a uniform non-zero electric field is the former but not the latter.

[6]Presumably if everyday life involved speeds $v \simeq c$ we would have a different intuition and none of this course would be surprising.

Consider a light signal emitted at a first event with coordinates (t_1, x_1, y_1, z_1) in a frame S, which is detected at a second event with coordinates (t_2, x_2, y_2, z_2). Since the signal travels at the speed of light

$$c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = 0 \ . \tag{33}$$

Given that the speed of light is the same in all inertial frames, we likewise have that the coordinates of these two events in S' satisfy

$$c^2(t_2' - t_1')^2 - (x_2' - x_1')^2 - (y_2' - y_1')^2 - (z_2' - z_1')^2 = 0 \ . \tag{34}$$

The interval

Motivated by Eqs. (33) and (35), for any two events we define the ***interval*** s between them in a frame S by

$$s = \left(c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2\right)^{1/2} \ . \tag{35}$$

For events infinitesimally close to each other, the interval is

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \ . \tag{36}$$

We can think of the interval as giving the "distance" between two events in a four dimensional space (ct, x, y, z). Compared to ordinary Euclidean geometry, there is the critical difference that the squares of the coordinate differences contribute with varying signs (the resulting geometry is called *pseudo-Euclidean*).

By the argument above, if $s = 0$ in one frame then it is 0 in all frames; such intervals are called ***light-like***.

Now consider a general infinitesimal invariant interval squared, with value ds^2 in S, which is not necessarily light-like. Since ds^2 is linear in dt^2 etc. we say that this is a second order infinitesimal. By the chain rule $dt = \frac{\partial t}{\partial t'}dt' + \frac{\partial t}{\partial x'}dx' + \ldots$ and so on, therefore the expression for ds^2 in terms of the quantities of a frame S' involves e.g. dt'^2. Because $ds^2 = 0$ if and only if $ds'^2 = 0$ (where $ds'^2 = c^2dt'^2 - dx'^2 - dy'^2 - dz'^2$), the only possible relation is [7]

$$ds^2 = a\, ds'^2 \ . \tag{37}$$

The coefficient a cannot depend on the space and time coordinates (this would violate homogeneity) or the direction of the relative velocity (this would violate isotropy), so the only thing it can depend on is the absolute value of the relative velocities of the two frames.

Suppose there are three reference frames S, S_1, S_2 with S_1 and S_2 having velocity relative to S of $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively. Then

$$ds^2 = a(|\mathbf{v}_1|)ds_1^2 \ , \qquad ds^2 = a(|\mathbf{v}_2|)ds_2^2 \ , \tag{38}$$

[7] Although this is intuitively plausible, I (and Landau & Lifshitz) am being a bit fast here and should prove this claim carefully. This can be done using methods from linear algebra.

and

$$ds_1^2 = a(|\mathbf{v}_{12}|)ds_2^2 \ , \tag{39}$$

where $\mathbf{v}_{12}$ is the velocity of S_2 relative to S_1. Therefore

$$\frac{a(|\mathbf{v}_2|)}{a(|\mathbf{v}_1|)} = a(|\mathbf{v}_{12}|) \ . \tag{40}$$

$|\mathbf{v}_{12}|$ depends on the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$, but this angle does not appear on the right-hand side of Eq. (40), hence $a(v)$ must be a constant. Moreover, Eq. (40) tells us that this constant is equal to 1.

Plugging $a(v) = 1$ into Eq. (37), we see that $ds^2 = ds'^2$ in general, not only when $ds^2 = 0$. Since this applies for infinitesimal intervals it applies for finite intervals too:

Invariance of the interval

The interval s between any two events is invariant, i.e. it takes the same value in all inertial reference frames:

$$(ct_2-ct_1)^2-(x_2-x_1)^2-(y_2-y_1)^2-(z_2-z_1)^2 = (ct_2'-ct_1')^2-(x_2'-x_1')^2-(y_2'-y_1')^2-(z_2'-z_1')^2 \ . \tag{41}$$

Eq. (41) tells us the form of Lorentz transformations: they are "rotations" in the four-dimensional space ct, x, y, z that leave the interval unchanged. I won't prove it, but every such rotation can be built up as a combination of rotations in the xy, xz, yz, tx, ty, and tz planes. The first three of these transform only the space coordinates and leave spatial distances unchanged, so they correspond to the usual spatial rotations, as in Galilean relativity (see Eq. (4)).

Instead, consider a rotation in the xt plane, which we anticipate will describe boosts between frames moving with relative velocity along the x-axis. Under this the y and z coordinates do not change. Therefore, by the invariance of the interval, the coordinates of any event in the original and rotated frame must satisfy

$$c^2t^2 - x^2 = c^2t'^2 - x'^2 \ . \tag{42}$$

The most general relation between the primed and unprimed coordinates for which Eq. (42) holds for all t' and x' is

$$\begin{aligned} ct &= ct' \cosh\psi + x' \sinh\psi \\ x &= ct' \sinh\psi + x' \cosh\psi \ , \end{aligned} \tag{43}$$

where $\psi \in (\infty, \infty)$ is an effective "angle" of rotation. The trigonometric functions that appear in normal rotations have been replaced by hyperbolic functions due to the minus sign in Eq. (42). Exercise: convince yourself that Eq. (43) is the most general solution to Eq. (42).

Our final task is to identify the relation between ψ and the relative velocity of two frames. The spatial origin O' of the S' system, $(t', x', y', z') = (t', 0, 0, 0)$, has coordinates in S given by

$$x_{O'} = ct' \sinh\psi \ , \qquad ct_{O'} = ct' \cosh\psi \ , \qquad y_{O'} = 0 \ , \qquad z_{O'} = 0 \ . \tag{44}$$

The frame S' is therefore moving relative to S in the x-direction. Its speed is $v = x_{O'}/t_{O'}$. Dividing the first two of the preceding equations, this is given by

$$\frac{v}{c} = \tanh\psi \ . \tag{45}$$

The parameter ψ is called the ***rapidity***. From Eq. (45),

$$\sinh\psi = \frac{v/c}{\sqrt{1-v^2/c^2}} \ , \qquad \cosh\psi = \frac{1}{\sqrt{1-v^2/c^2}} \ . \tag{46}$$

Substituting these expressions into Eq. (43), we obtain one of the most important results in the course

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} \gamma_v & \frac{v}{c}\gamma_v \\ \frac{v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix} \ , \qquad y = y', \qquad z = z' \ , \tag{47}$$

where $\gamma(v) := (1 - v^2/c^2)^{-1/2}$ and we use the notation $\gamma(v) \equiv \gamma_v$.

Eq. (47) is often rewritten as the transformation from the S coordinates to the S' coordinates (in practice both forms are useful). To do so, note that the inverse of the 2×2 matrix in Eq. (47) is obtained by replacing v with $-v$. This is clear physically (S is moving with velocity $-v$ relative to S') and it is straightforward to check algebraically (analogously to Eq. (3)):

$$\begin{pmatrix} \gamma_v & \frac{-v}{c}\gamma_v \\ \frac{-v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} \gamma_v & \frac{v}{c}\gamma_v \\ \frac{v}{c}\gamma_v & \gamma_v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ . \tag{48}$$

We therefore obtain:

The Lorentz boost

Consider some spacetime event. Its coordinates in a frame S' that is moving with velocity v in the x-direction relative to the frame S are given by

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma_v & \frac{-v}{c}\gamma_v \\ \frac{-v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \ , \qquad y' = y, \qquad z' = z \ . \tag{49}$$

This transformation is called a ***Lorentz boost*** in the x-direction.

It's common to write the explicit factors of v/c in the Lorentz boost as another parameter $\beta := v/c$, which takes values in $(-1, 1)$.

Sometimes people call Eq. (49) *the* Lorentz transformation, but in general a Lorentz transformation is any possible combination of rotating and boosting (with no shift in the space or time origin). Eq. (49) is the analogue of the Galilean relativity expression Eq. (1). We often denote the matrix implementing a particular Lorentz transformation Λ. E.g. this could be the 2×2 matrix in Eq. (49) or a 4×4 matrix acting on $(ct, x, y, z)^T$, depending on the context.

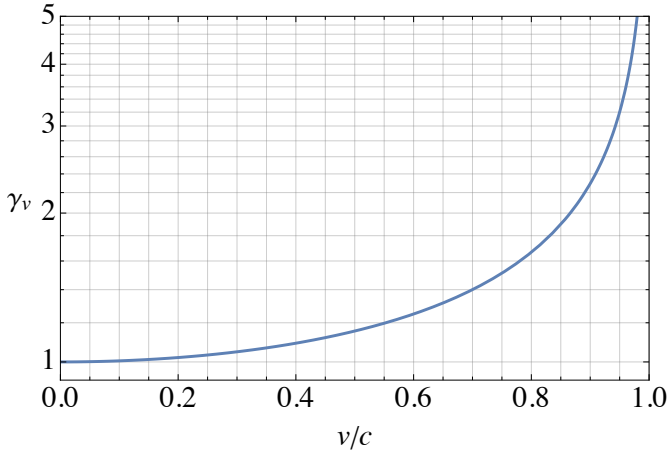


Figure 3: The function $\gamma_v := (1 - v^2/c^2)^{-1/2}$, plotted with a logarithmic scale on the vertical axis.

It is useful to be familiar with the graph of γ, plotted in Fig. 3. Note that γ is an even function, i.e. $\gamma_{-v} = \gamma_v$. For $v \ll c$ we have $\gamma_v \simeq 1 + \frac{1}{2}v^2/c^2$, while for $v \to c$ we have $\gamma_v \to \infty$. *Short exercise: calculate* $d\gamma/dv|_{v=0}$.

Eq. (49) has several crucial properties (it is a useful exercise to check these):

- It indeed leaves the interval invariant (i.e. I didn't make an algebra mistake).
- For $v > c$, γ_v is imaginary. This corresponds to the fact that speed of the frame S relative to S' cannot exceed the speed of light. Moreover, one cannot use Eq. (49) to boost to a frame moving at the speed of light.
- For $v/c \ll 1$ and $vx/(c^2) \ll t$, Eq. (49) reduces to the Galilean boost Eq. (1).
- As used already, the inverse transformation is obtained by replacing v with $-v$.
- The composition of two Lorentz boosts is itself a Lorentz boost (see the example below).
- Unlike a Galilean transformation, a Lorentz boost mixes the space coordinates into the time coordinate.
- The wave equation Eq. (18) is covariant under this transformation (with E_y unchanged).

Lorentz invariant and covariant mean the same thing as Galilean invariant and covariant, just with Galilean transformations replaced by Lorentz transformations. I sometimes won't write out "Lorentz" when it's obvious.

Example: Composition of Lorentz boosts

Consider three frames all moving relative to each other in the x-direction. Suppose S' is moving relative to S with velocity v, and S'' is moving relative to S' with velocity w. Since $c^2t^2 - x^2 = c^2t'^2 - x'^2$ and $c^2t'^2 - x'^2 = c^2t''^2 - x''^2$, we have $c^2t^2 - x^2 = c^2t''^2 - x''^2$ so the composition of the two transformations is also a Lorentz boost in the x-direction.

The explicit form of the transformation relating S to S'' is most easily obtained from the expression in terms of hyperbolic functions, Eq. (43). The composition of the transformation matrices is

$$\begin{pmatrix} \cosh\psi_w & \sinh\psi_w \\ \sinh\psi_w & \cosh\psi_w \end{pmatrix} \begin{pmatrix} \cosh\psi_v & \sinh\psi_v \\ \sinh\psi_v & \cosh\psi_v \end{pmatrix} = \begin{pmatrix} \cosh(\psi_v+\psi_w) & \sinh(\psi_v+\psi_w) \\ \sinh(\psi_v+\psi_w) & \cosh(\psi_v+\psi_w) \end{pmatrix} \quad , \tag{50}$$

using normal hyperbolic identities. The combined transformation corresponds to a single transformation with velocity u given by

$$u = c\tanh(\psi_v+\psi_w) = c\frac{\tanh\psi_v + \tanh\psi_w}{1+\tanh\psi_v\tanh\psi_w} = \frac{v+w}{1+vw/c^2} \in (-c,c) \ , \tag{51}$$

which is the velocity of S'' relative to S. The fact that ψ_v and ψ_w sum is just the analogue of how the angles sum for successive rotations (in contrast, velocities are not additive under successive boosts).

In passing, let me introduce a bit more terminology. The spacetime of special relativity is topologically $\mathbb{R}^4$, but with the addition of the invariant interval s as a measure of distance it is referred to as ***Minkowski space***. To stress the different ways that time and space contribute to s, spacetime is sometimes said to be $1+3$ rather than 4 dimensional.

The set of all combinations of rotations and Lorentz boosts is known as the proper-orthochronous Lorentz group (we have verified some of its group properties above). Including also space and time translations leads to the Poincaré group.

The designation "proper-orthochronous" is because I glossed over a subtlety. Reflections in time, which change the sign of time-like vectors (see below), are also part of the Lorentz group, as are spatial reflections that change the orientation of the coordinate axis. These are not continuously connected to the part of the Lorentz group containing the identity transformation that we are interested in. It turns out that parts of our best current theory of fundamental physics are invariant under spatial and time reflections but parts are not (namely, weak interactions).

Transformation of the separation between two events

Since Lorentz transformations are linear, the displacement vector between two events E_1 and E_2 transforms in the same way as the coordinate vectors themselves

$$\begin{pmatrix} ct_2' - ct_1' \\ x_2' - x_1' \end{pmatrix} = \Lambda \begin{pmatrix} ct_2 \\ x_2 \end{pmatrix} - \Lambda \begin{pmatrix} ct_1 \\ x_1 \end{pmatrix} = \Lambda \begin{pmatrix} ct_2 - ct_1 \\ x_2 - x_1 \end{pmatrix} , \tag{52}$$

where Λ is the 2×2 matrix in Eq. (49). (Note however, care is needed when considering translations in the space and time origins, which are not Lorentz transformations.)

Example: Lorentz transformations

An inertial frame S' is travelling at speed $0.4c$ in the positive x-direction relative to S. Two events in S' occur simultaneously 100 m apart. What is the separation in time of these events in S?

The difference between the coordinates of the events in S' is $(ct_2' - ct_1', x_2' - x_1') = (0, 100\,\mathrm{m})$, so

$$\begin{pmatrix} ct_2 - ct_1 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} \gamma_v & \frac{v}{c}\gamma_v \\ \frac{v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} 0 \\ 100\,\mathrm{m} \end{pmatrix} . \tag{53}$$

Therefore, the time separation in S is $t_2 - t_1 = 100\,\mathrm{m}\frac{v}{c^2}\gamma_v \simeq 43.6\mathrm{m/c} = 145\mathrm{ns}$ (with the event at larger x happening later).

Example: The end of universal time, version 2

Returning to the thought experiment of Fig. 2, we can now calculate the time when the signals arrive at B and C in the frame S'. To make the algebra simple let's choose the spacetime origin of two frames such that the signal is emitted at $(t, x) = (0, 0) = (t', x')$.

Let the distance from A to B and from A to C in the frame S be d. In the frame S the signal arrives these B and C at $(ct, x) = (d, -d)$ and $(ct, x) = (d, d)$ respectively.

Using Eq. (49), in S' the signal hence arrives at B at

$$(t', x') = \left(\gamma_v \frac{d}{c} \left(1 + \frac{v}{c} \right), -\gamma_v \left(d + v/c \right) \right) = \sqrt{\frac{1 + v/c}{1 - v/c}} \left(\frac{d}{c}, -d \right) . \tag{54}$$

Similarly, the signal arrives at C at

$$(t', x') = \sqrt{\frac{1 - v/c}{1 + v/c}} \left(\frac{d}{c}, -d \right) . \tag{55}$$

As expected, the signal reaches B and C at different times in the frame S', so simultaneity is not universal. In general two spatially separated events that are simultaneous in one frame

are not simultaneous in a boosted frame. Note that $x'/t' = c$ for the both the signals travelling to B and C, as must be the case.

In the limit $v \to c$, in the S' frame the light takes arbitrarily long to catch up to the point B since both are moving to the left at almost the speed of light. Conversely, the light reaches C almost immediately and travels almost no distance from its starting point $(t', x') = (0, 0)$.

Example: More Lorentz transformations

At $t = 0$ in the rest frame of the Earth, an alien civilisation in the Andromeda Galaxy ($d = 2.5$ million light-years away) votes to send a spaceship, dispatched immediately, to destroy the Earth. Also at $t = 0$, two people (Alice and Bob) are walking past each other at a speed of about 1ms^{-1} each. Alice happens to be walking towards Andromeda and Bob away from it. What is happening in Andromeda simultaneous to the event of Alice and Bob crossing in each of their reference frames?

Call the Earth's, Alice's, and Bob's frames S, S', S'' respectively. Choose the spacetime origins such that this at the event when Alice and Bob cross in all three frames. Then the vote, at event with coordinates in S of $(ct, x) = (0, d)$, has coordinates in S' of

$$\begin{pmatrix} ct' \\ d' \end{pmatrix} = \begin{pmatrix} \gamma_v & \frac{-v}{c}\gamma_v \\ \frac{-v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} 0 \\ d \end{pmatrix} \simeq \begin{pmatrix} -3\,\text{days} \\ d \end{pmatrix} \tag{56}$$

so the vote happened already a few days ago in Alice's frame when she meets Bob. Conversely, the time-coordinate of the vote event in Bob's frame is $ct'' = 3\,\text{days}$, so it has not yet happened when he meets Alice.

The slow speeds in this example are compensated by the large distances so that the coordinate transformation is not well approximated by the Galilean transformation $t = t' = t''$. This is why I was careful above that the Lorentz transformations reduce to Galilean transformations only when $v/c \ll 1$ *and* $vx/(c^2) \ll t$.

As will be clear in Section 9, there is no way that Alice and Bob can know about the vote when they meet, so Alice can't warn Bob about what has already happened in her frame and tell him to negotiate!

8 Digression: natural units

This Section is not essential for the present course, but it is indispensable for any physicist! Parts of the discussion here are inspired by Prof. R.L. Jaffe's very nice notes, which are available at: `https://stuff.mit.edu/afs/athena/course/8/8.06/spring05/handouts/units.pdf`

In Eq. (2), I wrote the Lorentz boost as acting on (ct, x) rather than (t, x), because doing so makes the expression for the transformation matrix nicely symmetrical. Similarly, from the expression for the interval, Eq. (35) we can see that ct and x enter in almost the same way (with the difference that the square of the time plays a privileged role because it doesn't come with a minus sign). The (almost) symmetry between space and time coordinates is somewhat obscured by the presence of c and it is also a hassle to keep track of these factors in expressions.

Moreover, the numerical value of c depends on the system of units used, and it's large value 3×10^8 in units of ms^{-1} is simply an artifact of metres and seconds being convenient for everyday life where speeds are small compared to the speed of light.

When doing special relativity, it is much nicer to adopt a system of units in which the numerical value of c is equal to 1 since this is what sets the natural scale by which to judge whether a speed is fast or slow: working in such units we can immediately see a speed $v = 0.01c$ is slow by the standards of relativity (e.g. γ_v is well approximated by the first few terms in its series expansion), whereas writing $v = 3 \times 10^6\,\mathrm{ms}^{-1}$ this is less clear.

Such a system of units would also treat time and space almost symmetrically in the interval. To clarify this, suppose that we choose to measure distances in the z-direction in inches and distances in the x- and y-directions in cm. Then the space part of the interval would be $(x_1 - x_2)^2 + (y_1 - y_2)^2 + b^2(z_2 - z_1)^2$ where $b = \mathrm{cm/inch}$ is a conversion factor. Of course carrying around such a conversion factor isn't a good idea and it would be much nicer to measure distances in the z-direction in cm as well, so that $b = 1$. The situation is much the same with c.

How do we implement such a system of units? Recall that in SI units all mechanical quantities are expressed as combinations of the kilogram, metre and second.[8] However, there's no fundamental reason why mass, length, and time must form the basis of our unit system. Let us instead choose velocity as one of our base units and take the units of velocity to be c. We could choose any two other linearly independent quantities as base units. In the context of fundamental physics the standard choice is to use momentum×length (later in your course you'll recognise this as the units of an action) measured in units of $\hbar$, and energy measured in units of eV. All physical quantities can then be expressed in terms of these three base units. Any quantity expressed, say, in SI units can be re-expressed in such units, by making use of the SI expressions for c, $\hbar$ and eV: $c \simeq 3 \times 10^8\,\mathrm{ms}^{-1}$ etc and their inverses, e.g. $1\,\mathrm{s} \simeq 1.52 \times 10^{15}\,\hbar\,\mathrm{eV}^{-1}$ etc. This system is called ***natural units***. The statement that c is equal to 1 in natural units is really no difference to saying a metre is equal to 1 in SI units.

We make a final, potentially confusing, step by dropping factors of c and $\hbar$ when we write quantities in natural units, so, for example, the mass of an electron is written as $m_e \simeq 0.5\,\mathrm{MeV}$. It is understood that the necessary factors of c and $\hbar$ to restore dimensional consistency are implicit (e.g. really $m_e \simeq 0.5\,\mathrm{MeV}/c^2$). Note that in this way many physically different quantities appear

[8] When dealing with electromagnetism, SI units introduce a fundamental unit of charge, the Coulomb, but this is not essential. Indeed, in so-called CGS units *all* physical quantities are written as combinations of centimetres, grammes and seconds, with the unit of charge defined by Coulomb's law.

to carry the same units (e.g. the units of length and time are both eV^{-1} and force and magnetic field are both eV^2), so you have to pay attention to the context to know how factors of c and $\hbar$ should be restored.

As mentioned, and anticipated by their name, natural units have the advantage of measuring quantities in terms of the natural scales set by relativity and quantum mechanics. In doing so, some of the basic equations reduce almost to identities $E = mc^2$, $\lambda_{\text{Compton}} = \hbar/(mc)$ etc. Moreover, simply by constructing quantities with different dimensions out of the scales relevant in a problem and c and $\hbar$ (essentially, just doing dimensional analysis), we can learn a remarkable amount about fundamental physics. If you did the exercise in red on page 2, you were basically doing such an analysis.

At this point, I would recommend reading Sections 4 and 5 of the notes by Jaffe linked above (some of the examples might not be familiar to you yet, but you should be able to follow the broad theme).

Finally, note that an extended version of natural units is possible with $c = \hbar = G_N = 1$ and dropped, which means that all physical quantities are dimensionless. This system is mostly used when working on quantum gravity.

Having said all that, for the rest of this course we will maintain factors of c.

9 Causality

Although events that are simultaneous in one inertial frame are not simultaneous in others, a Lorentz invariant notion of causality (that a cause leads to an effect) survives in special relativity. This is fortunate! It's hard to imagine how we could do physics without all observers agreeing on causality.

Working in an inertial frame S, imagine a spacetime diagram with the ct axis vertical and the space axes perpendicular to this. An example is plotted in Fig. 4, with only one space dimension shown for simplicity. We divide spacetime into three regions separated by the surface $x^2 + y^2 + z^2 = c^2t^2$, which is two cones connected at the origin (in Fig. 4 this reduces to two triangles, with surface $x = \pm ct$). The cone in the $t \geq 0$ region is the set of events that a light signal emitted in all directions from an event O at $(0, 0, 0, 0)$ would travel through and is known as the ***future light-cone*** of O. The cone in the $t < 0$ region is the set of all events from which light signals would arrive at O and is called the ***past light-cone*** of O.

Suppose that at $t = 0$ a particle of non-zero mass is at O. As time goes on, the particle will trace out a path, called its ***worldline***. Because the speed of the particle is strictly less than the speed of light, the worldline necessarily lies inside the future light-cone. Similarly its worldline must have come from inside the past light cone. The particle cannot reach or have come from

outside the light cones. We can construct light cones around every point on the worldline, and the particle must go from the past light cone into the future light cone of all of these.

The division of spacetime into regions inside and outside any given event's light-cones is Lorentz invariant. For example, if an event E is outside O's future light cone in S then this is true in all frames. This can be seen by the following classification.

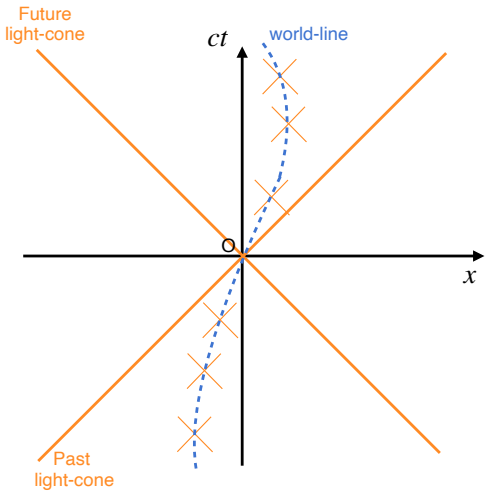


Figure 4: The past and future light cones of an event O. A particle of non-zero mass passing through O can only travel inside of the future light cone and must have come from inside the past light cone. Moreover, the same must be true for the light cones at every event on its worldline (some of which are drawn). Events within, or on the boundary of, O's future light cone can be affected by something that happens at O. Similarly, a physical system at O can only have been affected by things that happened inside O's past light cone.

Classification of invariant intervals

Consider two events $E_1 = (ct_1, x_1, y_1, z_1)$ and $E_2 = (ct_2, x_2, y_2, z_2)$, separated by invariant interval s.

- E_2 is inside the future or past light cone of E_1 if

$$s^2 = c^2(t_1 - t_2)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2 - (z_1 - z_2)^2 > 0 \ , \tag{57}$$

  in which case, we say E_2 and E_1 are ***time-like*** separated.

- E_2 is on the surface of the future or past light cone of E_1 if

$$s^2 = c^2(t_1 - t_2)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2 - (z_1 - z_2)^2 = 0 \ , \tag{58}$$

  in which case E_2 and E_1 are ***light-like*** separated.

- E_2 is outside E_1's light cones if

$$s^2 = c^2(t_1 - t_2)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2 - (z_1 - z_2)^2 < 0 \ , \tag{59}$$

  in which case E_2 and E_1 are ***space-like*** separated.

We similarly say that the displacement vector between E_2 and E_1, $(ct_2 - ct_1, x_2 - x_1, y_2 - y_1, z_2 - z_1)$, is time-like, light-like or space-like, respectively.

Since the interval is Lorentz invariant, this classification is Lorentz invariant.

Moreover, for time-like separated events the sign of $t_2 - t_1$ is unchanged by any Lorentz transformation. *Proof:* The transformed coordinates are continuous functions of the relative velocity of the frames, so if there was a transformation that changed the sign of $t_2 - t_1$ there would also be a transformation for which $t_2 - t_1 = 0$. However, this transformation would have interval $s^2 \leq 0$, which is a contradiction. By studying Eq. (2) it can be shown that the same result is also true for light-like separated events.

We hence obtain a Lorentz invariant notion of causality.

Causality in special relativity

Two events with a space-like separation cannot be causally connected because physical interactions propagate from one point to another with speed no greater than that of light.

Events with time-like or light-like separation can be causally connected, and the time ordering of such events is the same in all inertial frames.

In other words, a physical system at O can only have been affected by things that happen within, or on the boundary of, O's past light-cone. Likewise, it can only affect things inside, or

on the edge of, O's future light cone. This is a more restrictive version of causality than occurs in Galilean relativity, for which information can travel infinitely fast.

The following facts are useful to know (some of these are proved on the first tutorial sheet, and the others can be proved similarly).

- For time-like separated events there always exists an inertial frame in which the two events occur at the same point in space.
- For any particle of non-zero mass there always exists a reference frame in which it is instantaneously at rest. (Consider two nearby events on its worldline. Since these are time-like separated there is a frame in which they are at the same point in space.)
- For any space-like separated events there exists an inertial frame in which they occur simultaneously and the time-ordering can also always be reversed by going to a different frame.

Example: Finding particular frames
Find a frame S' in which two time-like separated events, E_1 and E_2 with coordinates in S of (ct_1, x_1) and (ct_2, x_2), occur at the same point in space.

The difference between the coordinates of the two events in S' must be $(ct'_2 - ct'_1, 0)$. Hence, we need find a $v < c$ such that

$$\begin{pmatrix} \gamma_v & \frac{-v}{c}\gamma_v \\ \frac{-v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} ct_2 - ct_1 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} ct'_2 - ct'_1 \\ 0 \end{pmatrix} . \tag{60}$$

i.e. we must solve

$$-\frac{v}{c}\gamma_v(ct_2 - ct_1) + \gamma_v(x_2 - x_1) = 0 , \tag{61}$$

for v. This can be done easily, $v/c = (x_2 - x_1)/(ct_2 - ct_1)$.

Example: Invariant intervals
Some examples of events with different types of separation are plotted in Fig. 5.

Extending our terminology a bit, we say that a worldline is time-like if the infinitesimal interval $ds^2 > 0$ everywhere along it. As above, the world line of a massive particle is time-like because $dx^2 + dy^2 + dz^2 = v^2dt^2 < c^2dt^2$. Similarly a worldline is light-like if $ds^2 = 0$ all along it, corresponding to the worldline of a photon (or other massless particle). Finally, a worldline is space-like if $ds^2 < 0$ all along it.

[10] *Caution: There isn't actually a single Lorentz transformation that simultaneously has the plotted effect on all the intervals.*

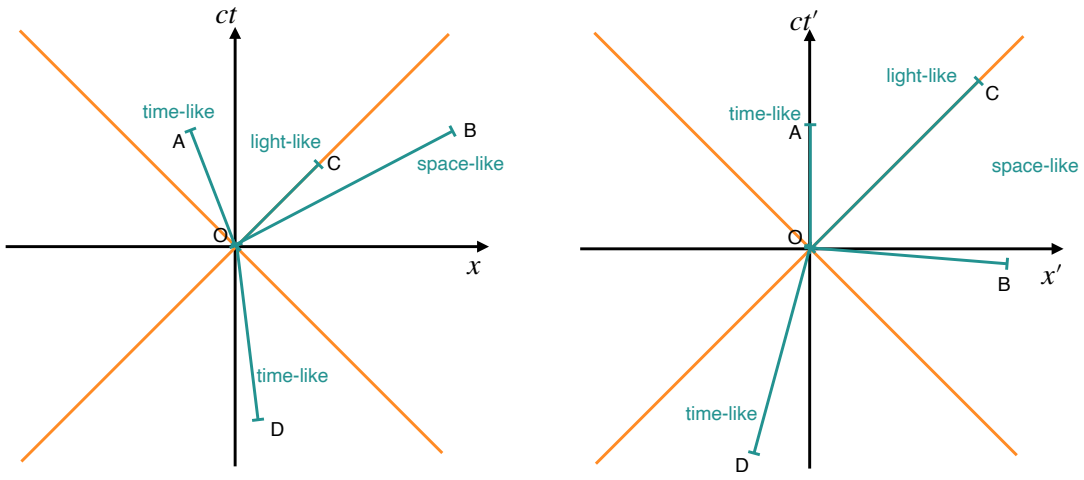


Figure 5: **Left:** The invariant interval between two events can be classified as time-like (as for OA and OD) or space-like (e.g. OB) or light-like (OC). **Right:** Examples of what can happen on Lorentz transforming from the inertial frame in the left panel to a new inertial frame (primed coordinates). There always exists a frame where two time-like separated events occur at the same point in space, e.g. OA. The time ordering of space-like separated events can reverse, e.g. OB.[10]

A similar concept of causality persists in quantum field theory (the combination of special relativity and quantum mechanics) but with interesting subtleties. A nice discussion has been given by Prof. Cliff Burgess (slightly edited):

> *In a quantum theory, if you measure exactly where a particle is then you know nothing about how fast it is moving, and in particular there is a non-zero probability for it to be moving faster than the speed of light. Suppose A and B are space-like separated events and you think A is earlier than B. Then you might see that there is a non-zero probability for a particle to carry charge Q from A to B. For this to be consistent with causality, I (in another inertial frame) have to be able to think something went from B to A carrying exactly charge $-Q$. Moreover, the quantum amplitude for this must be exactly the same as for the process you saw. The particle that does that for me is the antiparticle of the one you saw.*

Note that, importantly, there is still no way to transmit information from A to B. There is nothing you can do at A that will affect a measurement at B (put mathematically, all space-like separated operators commute).

10 Spacetime diagrams

It can be useful to show the coordinate axes for two inertial reference frames related by a Lorentz boost on a single plot, analogously to the Galilean transformation plot of Fig. 1.

From Eq. (49), lines of constant t' (with the particular case $t' = 0$ being the x' axis) correspond to $ct' = \gamma_v ct + \frac{v}{c}\gamma_v x$, i.e.

$$ct = \frac{v}{c}x + \frac{ct'}{\gamma} \ . \tag{62}$$

Similarly, lines of constant x' are $x' = -v\gamma_v t + \gamma_v x$, so

$$ct = \frac{c}{v}x - \frac{c}{v}\frac{x'}{\gamma} \ . \tag{63}$$

Note that Eq. (63) can be rewritten as $x = \frac{v}{c}(ct) + \frac{x'}{\gamma}$, which shows the symmetry between the x' and ct' axes. From these expressions we see that the x' and ct' axes are at an angle $\theta = \arctan(v/c)$ to the x and ct axes. They both point into the first quadrant if $v > 0$ and point into the second and fourth quadrants if $v < 0$.

An example spacetime diagram with S' moving at velocity $v > 0$ in the S frame is shown in Fig. 6 left, in particular with $v/c \simeq 0.6$ so $\gamma_v \simeq 1.25$. Note that, from Eq. (62) and (63), the "tick marks" on the x' and ct' axes are further apart than on the x and ct axes. For example, the event with coordinates in S' of $(ct', x') = (0, 1)$ corresponds to $(ct, x) = \gamma_v(v/c, 1)$. This is unlike the Galilean case. An equivalent way to locate the tick marks on the S' axes is using the invariant interval: the point $(ct', x') = (0, 1)$ lies at the point where the x' axis intersects the hyperbola $c^2t^2 - x^2 = -1$.

Given an event such as the red star, we can read off its coordinate in the two frames. In the S frame, we simply follow horizontal and vertical lines as usual, and in the S' frame we follow the lines of constant x' and ct' to the S' axes. The faint orange lines show the worldlines of light emitted from the origin $(ct, x) = (0, 0) = (ct', x')$; this is the line $ct = \pm x$ which coincides with the line $ct' = \pm x'$ because light travels at the same speed in both frames.

Finally, note that in the case of a transformation with $v < 0$, the S' axes are aligned as in Fig. 6 right. As an example, the worldline of a particle at a fixed space position in the S' frame is plotted.

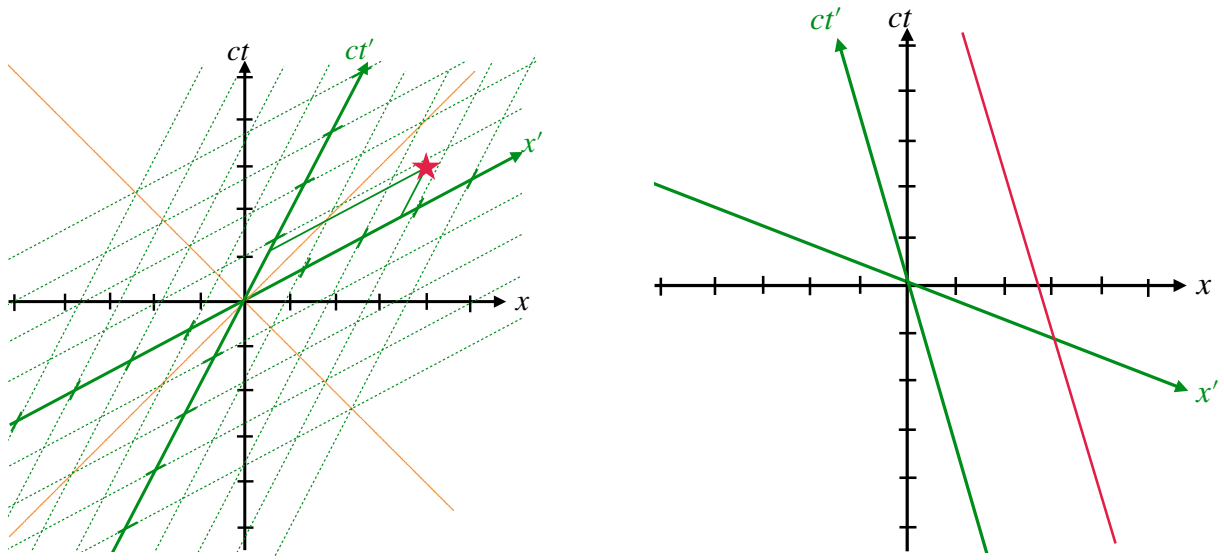


Figure 6: **Left:** spacetime diagram showing the coordinates in two inertial frames S and S' related by a Lorentz boost with $v > 0$ (v being the velocity of S' in the S frame). The event with a red star has coordinates $(ct, x) = (3, 4)$ and $(ct', x') \simeq (0.9, 2.8)$ in the two frames. **Right:** The analogous plot for $v < 0$. In this case a particle of fixed position in the S' frame, with its worldline shown in red, has an x coordinate that decreases as t increases.

11 Time dilation and length contraction

Two interesting implications of special relativity are time dilation ("moving clocks run slow") and length contraction ("moving objects are shortened").[11] Both of these can be derived from the Lorentz transformation.

[11] A more precise statement of time dilation is that "viewed from an inertial frame a moving clock runs slow". For example, later we will discuss the twin paradox, and in this the twin that goes out and back (and has not remained in a single inertial frame) believes that a clock held by the one that remains on Earth has run fast not slow.

Time dilation

Consider a clock that is stationary and with time interval between its ticks of τ in a frame S. This means that the spacetime separation between two successive tick events E_1 and E_2 in S is $(ct_2, x_2) - (ct_1, x_1) = (c\tau, 0)$. What is the separation in time between the tick events measured in a frame S' moving relative to S with velocity $\mathbf{v}$ in the x-direction?

Using Eq. (49), the difference between the tick events in S' is

$$\begin{pmatrix} ct_2' - ct_1' \\ x_2' - x_1' \end{pmatrix} = \begin{pmatrix} \gamma_v & \frac{-v}{c}\gamma_v \\ \frac{-v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} c\tau \\ 0 \end{pmatrix} \ . \tag{64}$$

Therefore

$$t_2' - t_1' = \gamma_v \tau \ , \tag{65}$$

so the time interval between ticks is longer in S' than in the clock's rest frame, i.e. to an observer in S' the clock runs slow. This is ***time dilation***.

Since the time-coordinate is not affected by spatial rotations, Eq. (65) holds for any frame moving with speed v relative to S.

The same argument holds for any process be it clocks, the decay of elementary particles, or the beating of a human heart. It is not that the physical properties of the clock change, instead time itself runs differently.

An even slicker derivation of the time dilation can be obtained from the invariance of the interval. The interval between tick events in the rest frame of the clock is

$$s^2 = c^2\tau^2 \ . \tag{66}$$

Meanwhile, the same interval evaluated in S' is

$$\begin{aligned} s^2 &= c^2(t_2' - t_1')^2 - (x_2' - x_1')^2 \\ &= c^2(t_2' - t_1')^2 - v^2(t_2' - t_1')^2 \ . \end{aligned} \tag{67}$$

Equating Eqs. (66) and (67)

$$c^2\tau^2 = c^2(t_2' - t_1')^2 - v^2(t_2' - t_1')^2 \ , \tag{68}$$

and solving for $t_2' - t_1'$ leads to Eq. (65).

At this stage we introduce some useful terminology. Consider two time-like separated events, and work in the inertial frame in which they are at the same location in space. The time between the events in this frame is called the ***proper time*** between them. For example, in Eq. (65), the proper time between the clock ticks is τ. Indeed, in general the proper time is exactly the time interval measured by a clock in uniform motion that passes through the two spacetime events.

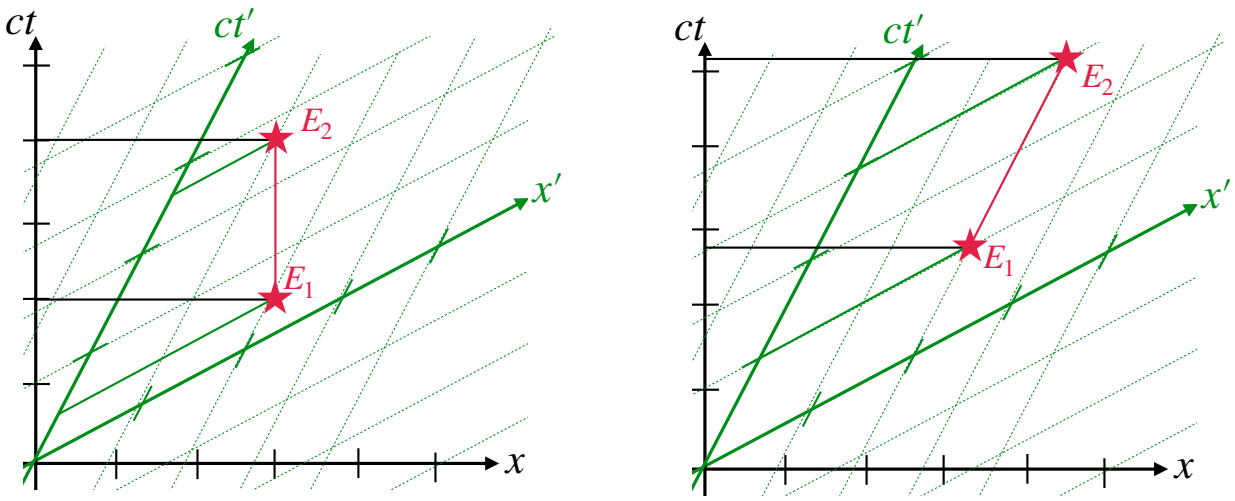


Figure 7: Time dilation. **Left:** Two events E_1 and E_2 at the same point in space in S separated by proper time $ct_2 - ct_1 = 2$ are separated in time by $ct'_2 - ct'_1 > 2$ in S'. **Right:** The opposite arrangement. Two events at the same point in space in S' separated by proper time $ct'_2 - ct'_1 = 2$ are separated in time by $ct_2 - ct_1 > 2$ in S.

Given the coordinates of two events in a frame S', analogously to Eq. (68) the proper time τ can be calculated from

$$c^2\tau^2 = c^2(t'_2 - t'_1)^2 - (x'_2 - x'_1)^2 \ . \tag{69}$$

The proper time between space-like or light-like separated events is not defined.

We can check our calculation of time dilation using spacetime diagrams. In Fig. 7 left we show the setup described above with two events E_1 and E_2 at the same point in space in S and hence at different locations in S'. The time-interval between them in S is $ct_2 - ct_1 = 2$ and $ct'_2 - ct'_1 \simeq 2.5$ in S' (allowing for my imperfect sketching).

In Fig. 7 right we show the opposite situation in which the two events are at the same location in S' and are separated by a proper time $ct'_2 - ct'_1 = 2$. In this case, the time-separation in S is $ct_2 - ct_1 \simeq 2.5$. This is as expected: from the perspective of S a clock passing through E_1 and E_2 is moving so must run slow.

In general, time dilation (and length contraction) is reciprocal. Two inertial observers in frames related by a boost both see clocks that are moving in their frame as ticking slowly compared to clocks that are at rest in their own.

To derive length contraction we first need a careful definition of an object's length (this is the same as used for Galilean relativity in Section 3, but let me repeat it here). Working in a frame S consider two events, both at the same time t, one at either end of the object. The spatial distance

between these events is the object's length in S. As we will see, the length of an object is not Lorentz invariant.

Proper length

The ***proper length*** of an object is its length measured in an inertial frame in which it is at rest. Note that in the rest frame of the object the two events used to define its length don't need to be simultaneous because the object doesn't move (in any other frame this is essential).

Length contraction

Suppose an object of proper length ℓ_0 is stationary and aligned along the x-axis in a frame S. What is its length ℓ' in a frame S' that is moving at velocity v relative to S in the x-direction?

Consider two events at either end of the object with coordinates in S' of (t', x'_1) and (t', x'_2) (note these are simultaneous in S'). The x-coordinates of these events in S are given by Eq. (47):

$$x_1 = \frac{v}{c}\gamma_v ct' + \gamma_v x'_1 \ , \qquad x_2 = \frac{v}{c}\gamma_v ct' + \gamma_v x'_2 \ , \tag{70}$$

so

$$x_2 - x_1 = \ell_0 = \gamma_v (x'_2 - x'_1) \ , \tag{71}$$

where we used that $x_2 - x_1 = \ell_0$ even though these events are not simultaneous in S. Hence,

$$\begin{aligned} \ell' &= x'_2 - x'_1 \\ &= \frac{\ell_0}{\gamma_v} \ . \end{aligned} \tag{72}$$

Since $\gamma \geq 1$ the length of the object in S' is smaller than its proper length. This is ***length contraction***.

The calculation above can again be explained with a spacetime diagram, plotted in Fig. 7 left. An object, stationary in S, traces out the worldsheet (the 2d surface analogous to a worldline) given by the red rectangle. The object's length in S is the distance Δx between the events E_2 and E_1 (or between E_4 and E_3). Its length in S' is the distance $\Delta x'$ between the events E_4 and E_3 (note these are simultaneous in S'). Reading off from the axis, the length of the object in S is 2 and in S' is about 1.6 in the units of the plot.

To check I haven't made a mistake, let's make sure that an object that is stationary in S' has a contracted length in S, as must be the case. This arrangement is plotted in Fig. 8 right. In the S' coordinates, the object has length 2 (measured again by e.g. the E_3 and E_4 events). Meanwhile the length in S measured by the E_1 and E_2 events is indeed contracted to about 1.6, as expected.

Distances orthogonal to the direction of a boost do not undergo length contraction since these coordinates have trivial transformation, e.g. $y' = y$, $z' = z$ for a boost in the x direction.

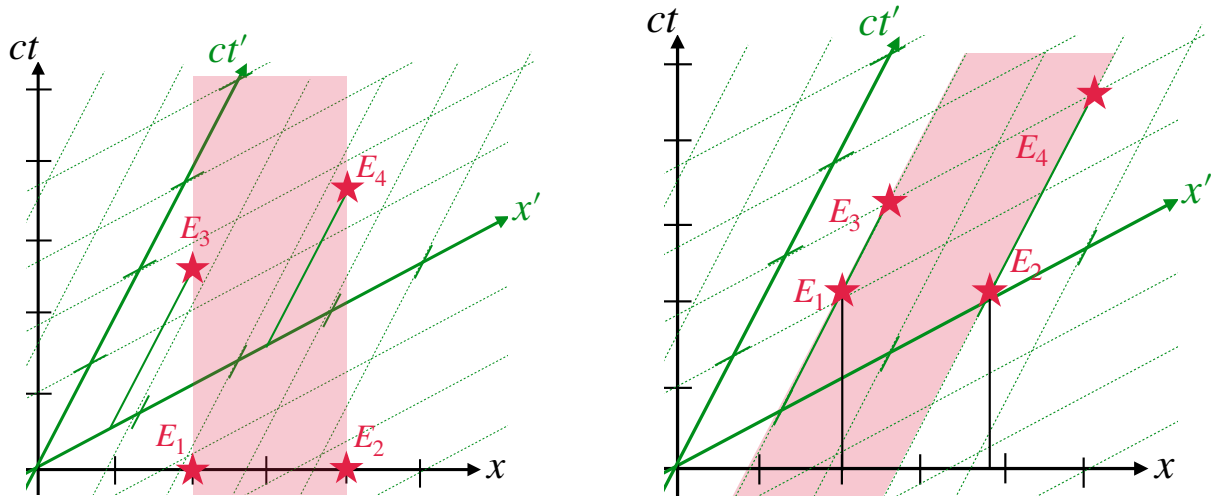


Figure 8: Length contraction **Left:** An object that is stationary in S sweeps out the red world-sheet. Its length in S (its proper length) is measured by the x-coordinates of the events E_1 and E_2, giving a length of 2. In S' its length is measured by x'-coordinates of the events E_3 and E_4, and is less than 2 due to length contraction. **Right:** The analogous plot for an object that is stationary in S'. The length in S is again measured by the events E_1 and E_2, and is contracted relative to the proper length (measured in S' via E_3 and E_4).

Let us also note that one can justify time dilation and length contraction by analysing particular physical setups. A brief outline of how this is done is given in Appendix A, and further details can be found in e.g. the book by French.

It might seem strange that time is dilated but length is contracted, which appear to be opposite effects, when the Lorentz transformations are symmetric in time and space. The point is that time plays a different role in how we experience the Universe compared to space, so this affects the measurements that we are interested in. Suppose that a clock and extended object are both stationary in S. In deriving time dilation we are interested in two events on the clock's worldline, which are separated in time and space in S'. However, in deriving length contraction we are interested in two events that in S' are separated in space but not in time. This difference breaks the symmetry between space and time.

Time dilation and length contraction are verified daily in high-energy physics laboratories as the next two examples illustrate.

Example: Pion beams

At the Fermi National Laboratory a beam of charged pions is produced with speeds of about $(1-10^{-6})c$. The lifetime of a charged pion at rest is $t_{\text{decay}} \simeq 3\times10^{-8}\,\text{s}$. Estimate the fraction of the beam of pions is lost to decays after travelling a distance of 300 m (in the lab frame) down a beam pipe?

The charged pions have $\gamma \simeq 700$. The time in the lab frame for them to travel 300 m is about 10^{-6} s, which corresponds to a proper time for the pions of $\tau \simeq 1.5 \times 10^{-9}$ s. The fraction of pions surviving is $e^{-\tau/t_{\rm decay}} \simeq 0.95$. If it wasn't for time dilation only a fraction $\sim 10^{-15}$ of pions would survive.

Note that in the rest frame of the pion the same result is obtained, but it is interpreted as length contraction (of course the two ways of doing the calculation are equivalent). In the pion rest frame the 300 m beam pipe is moving towards them and its length is contracted to $300\,{\rm m}/\gamma$. In this frame, the time interval between the pions being at the start and end of the beam pipe is again $\tau \simeq 1.5 \times 10^{-9}$ s. To clarify the perspective in the two frames even further, you could Lorentz transform the events with lab frame coordinates $(ct, x) = (0, 0)$ and $(10^{-6}c\,{\rm s}, 300\,{\rm m})$ to the pion rest frame.

Another well-known and very similar example is the decay of muons in the atmosphere. Just like the Fermi Lab pions, on average these travel much further than they otherwise would before decaying due to time dilation. There are also experiments on Earth that are carried out with muons. In fact, these experiments are only possible because time dilation increases the lab frame lifetime of fast moving muons enough for them to survive for a while.

Example: Length contraction and heavy ions
Sometimes the LHC is used to collide symmetric beams of lead nuclei with γ factors of about 2700 in the lab frame. Estimate the time for which the two colliding lead nuclei overlap.

The radius of a hydrogen nucleus is about fm so we estimate the radius of lead nuclei at rest to be about $200^{1/3}\,{\rm fm} \simeq 10\,{\rm fm}$. Taking into account the length contraction of the nuclei, in the lab frame they overlap for

$$t \simeq \frac{10\,{\rm fm}}{\gamma c} \simeq 10^{-27}\,{\rm s}\ . \tag{73}$$

In the lab frame the ions appear highly compressed into disks, which enhances their density by a factor of γ.

Example: Michelson–Morley II
Although the null result of the Michelson–Morley experiment is obvious when working in the rest frame of the experiment, as an exercise let's check the calculation in a frame in which the apparatus is moving at speed v in the direction of one of the arms. This could be, for example, the rest frame of the Sun. As in Eq. (28), the time for the light to travel to and back along the perpendicular arm is given by

$$t_2 = \frac{2L_2}{c}\gamma_v\ , \tag{74}$$

where L_2 is now the proper length of the arm. However, due to length contraction, the time for

the light to travel to and back along the parallel arm is modified from Eq. (25) to

$$t_1 = \frac{2L_1}{c\gamma_v}\gamma_v^2 = \frac{2L_1}{c}\gamma_v \tag{75}$$

where L_1 is the proper length of the arm. Since the preceding two expressions take the same form, there is no phase shift when the experiment is rotated.

The next example is less realistic.

Example: Space travel
A potentially inhabitable planet is at a distance of 20 light years from Earth. Suppose that a futuristic rocket is capable of travelling at $0.9c$

- According to observers on Earth, how long will it take the rocket to reach the star?
  – In the reference frame of the Earth the time is $20/0.9\,\text{yr} \simeq 22.2$ yr.
- According to passengers on the rocket, how long will it take them to reach the star?
  – The elapsed passenger proper time is shorter than the time elapsed in the Earth's frame by a factor of $\gamma \simeq 2.3$, so the journey takes $9.7\,\text{yr}$ according to them (equivalently, they see the distance to the star Lorentz contracted).
- When should an observer on Earth look through a (totally unrealistic) telescope to see the spaceship arrive at the star.
  – In the Earth's reference frame it takes 22.2 years for the observers to arrive and 20 years for the light to travel back to Earth, so they should wait $42.2\,\text{yr}$.

Effects that are similar to time dilation but more dramatic occur in general relativity. For example, to an observer far from a black hole it takes infinite time for an object falling into the black hole to reach the event horizon although the proper time for this is finite.

12 Proper time

So far we have defined the proper time between two time-like separated events as the time elapsed on a clock undergoing uniform motion between them. It is useful to generalise the definition of proper time to apply also to a physical system, such as a particle, that might be changing velocity (i.e. accelerating) and therefore changing inertial reference frames.

Suppose the particle is moving with instantaneous velocity $\mathbf{u}$ in some inertial frame S'. In a time interval dt' its position changes by $d\mathbf{x}' = \mathbf{u}dt'$ and hence the square of the corresponding

infinitesimal invariant interval is

$$ds^2 = c^2 dt'^2 - |d\mathbf{x}'|^2 = c^2 dt'^2 \left(1 - \frac{|\mathbf{u}|^2}{c^2}\right) \ . \tag{76}$$

In the inertial frame S in which the system is instantaneously at rest $dt = d\tau$ and $d\mathbf{x} = 0$. Thus the invariant interval is $ds = cd\tau$ and so

$$d\tau = dt' \sqrt{1 - \frac{|\mathbf{u}|^2}{c^2}} = \frac{1}{\gamma_u} dt' \ . \tag{77}$$

Integrating, we obtain:

Proper time

The proper time τ along a particular time-like worldline connecting any two events, A and B, is given by

$$\tau = \frac{1}{c} \int_A^B ds = \int_{t'_A}^{t'_B} \frac{1}{\gamma(t')} dt' \tag{78}$$

where $\gamma(t')$ is the value of γ calculated from the instantaneous speed $u(t')$. Since $d\tau = ds/c$ it is Lorentz invariant, τ is too and the right-hand side of Eq. (78) can be evaluated in any inertial frame.

The proper time along a worldline gives the time interval measured by a clock moving along it. The previous definition for two events is equivalent to Eq. (78) in the special case that the worldline is a straight line, corresponding to constant velocity (if you're asked the proper time between events with no worldline specified this is what is meant).

Eq. (78) again shows the phenomenon of time dilation. The elapsed proper time is less than the time measured in a frame in which a clock is moving. This is made more general by the following result.

Claim: Consider two time-like separated events E_1 and E_2. Out of all time-like worldlines between them, a straight line leads to the greatest elapsed proper time. Such a worldline corresponds to an object travelling at constant velocity between the two events. Sometimes this is called the principle of maximal ageing.

Proof: Since proper time is Lorentz invariant we can work in any inertial frame. Pick the frame S' in which the two events are at the same point in space. The straight-line worldline is then just the worldline of an object at rest, so has $\gamma(t') = 1$ for all t' and $\Delta\tau = \int_{t'_A}^{t'_b} dt'$. Now consider any other worldline. Along this, the instantaneous velocity is non-zero for some finite range of t'; $\gamma(t') > 1$ at these times and $\gamma(t') \geq 1$ for all t'. Hence, for such a worldline, the right-hand side of Eq. (78) $\int_{t'_A}^{t'_b} \frac{1}{\gamma(t')} dt' < \int_{t'_A}^{t'_b} dt'$.[12]

[12] If we didn't work in this particular frame we could derive the same result by using that γ is an increasing function with positive second derivative or with the calculus of variations.

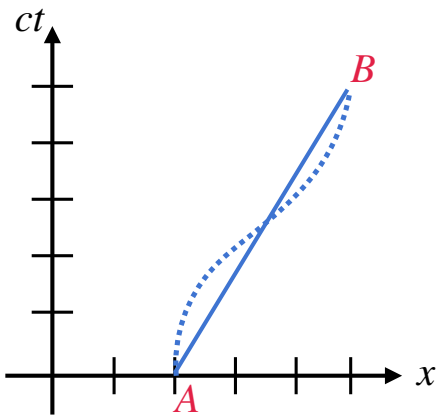


Figure 9: Two worldlines between spacetime events A and B. The solid line is the worldline of an object moving with constant velocity (in this case $\frac{3}{5}c$) while the dashed line is an object that accelerates. The proper time along the straight worldline is necessarily greater than along the curved one. Due to time dilation the proper time along the straight line τ_1 satisfies $c\tau_1 < 5$ (in the units of the plot) and the proper time along the curved worldline τ_2 is even smaller $\tau_2 < \tau_1$.

An example illustrating this result is shown in Fig. 9.

As we've seen in special relativity the trajectory of maximal ageing coincides with unaccelerated trajectory, i.e. the solution of the equations of motion in the absence of forces. In general relativity, this remains true. In this context, it is actually a useful way to calculate the motion of a body under gravity (thinking of gravity as a consequence of spacetime curvature rather than a force).

Example: Proper time

Consider two spacetime events E_1 and E_2 with coordinates in a frame S of $(ct, x) = (0, 0)$ and $(12, 5)$. Calculate the time interval measured by a clock a that travels in uniform motion between these events. Calculate the time interval measured by a clock b that first travels to reach $(ct, x) = (6, 5)$ and then remains stationary in the S' frame until E_2.

We could transform to the frame in which E_1 and E_2 are at the same spatial location, but instead let's work in S directly.

The clock that travels along a straight worldline between these events has speed $v = \frac{5}{12}c$, so $\gamma \simeq 1.1$ and hence has proper time $\tau_a \simeq 10.9$ (note the proper time is less than 12 as expected from time dilation).

Meanwhile, the other clock travels at $v = \frac{5}{6}c$, so $\gamma \simeq 1.8$ for $\Delta t = 6$ and is then stationary. Therefore, this worldline has proper time $\tau_b \simeq 6/1.8 + 6 \simeq 9.1$.

Example: Hafele–Keating experiment

An interesting, albeit given the previous examples perhaps not totally essential, test of time dilation has been carried with macroscopic clocks. The clocks were set in motion relative to the Earth by putting them in airplanes. The precision of the clocks was sufficient to overcome the relatively small speeds involved.

Four clocks were flown around the world, once in an eastward and once in a westward direction. During the journeys logs were kept of the aircrafts' location and ground speed so the integral in Eq. (78) could be calculated. Before and after each journey the clocks were compared with identical clocks at the U.S. Naval Observatory. With allowance for the Earth's rotation and the gravitational "redshift" of general relativity, the average observed and predicted time differences compared to the clock on the ground were (-59 ± 10) ns and (-40 ± 23) ns respectively for the eastward trips and (273 ± 7) ns and (275 ± 21) ns for the westward trips (the kinematic effect of special relativity is comparable to the general relativistic effect).

In the inertial frame following the centre of the Earth (ignoring acceleration due to the Sun) the clocks on the ground move due to the Earth's rotation. While travelling eastward on an airplane the clocks move even faster so have more time dilation, whereas while moving westward the clocks move slower so have less time dilation.

Sometime this year you will learn about the Lagrangian approach to Newtonian mechanics, which says that particles travel along trajectories that extremise a quantity called the "action". In the relativistic mechanics version of this, the action for a particle with no interactions is simply $-mc\int ds$, which is consistent with the observation above that this quantity is extremised by straight worldlines. Once there are e.g. electromagnetic fields present the action is more complicated. In string theory a similar analysis of the "worldsheet" rather than "worldline" leads to what is called the Nambu-Goto action.

13 Addition of velocities

The Lorentz transformations allow us to relate the coordinates of events in one inertial reference frame to any other. Now we show how to relate the velocity of a particle in one inertial reference frame to another.

To start with we consider systems in which all velocities are in the same direction. Suppose that a particle has velocity w' in the x-direction in a frame S' that itself is moving relative to a frame S with velocity v in the x-direction. What is the velocity w of the particle in the frame S?

We actually already know the answer to this question from Eq. (51). This shows that the effect of two successive Lorentz boosts with speeds w' and v in the x-direction is the same as a

single boost of velocity

$$\frac{v + w'}{1 + vw'/c^2} \ . \tag{79}$$

The rest frame of the particle is related to the frame S by successive boosts of w' and v and hence is related to the frame S by a single boost given by Eq. (79). This is therefore the velocity of the particle in S.

For practice, let's derive the same result another way. The velocity of the particle in S' is $w' = dx'/dt'$ (where x' and t' are the coordinates that the particle's worldline passes through) and in the S frame is dx/dt. From Eq. (47), we have

$$\begin{aligned} dt &= \gamma_v(dt' + \frac{v}{c^2}dx') = \gamma_v(1 + \frac{v}{c^2}\frac{dx'}{dt'})dt' \\ dx &= \gamma_v(vdt' + dx') = \gamma_v(v + \frac{dx'}{dt'})dt'. \end{aligned} \tag{80}$$

Dividing the second by the first equation,

$$w = \frac{dx}{dt} = \frac{v + w'}{1 + \frac{vw'}{c^2}} \ , \tag{81}$$

which is the same result as above.

It's worth noting a few features of Eq. (81)

- In the limit $v/c \ll 1$ and $w'/c \ll 1$ we obtain $w \simeq v + w'$, which is simply the expected Galilean velocity addition law.
- A massive particle with speed less than the speed of light in S' also has speed less than the speed of light in S. Mathematically, $|w'| < c$ implies $|w| < c$. To prove this, fix v (with $|v| < c$) and consider w as a function of w'. Then

  $$\frac{dw}{dw'} = \frac{(1 - v^2/c^2)}{(1 + vw'/c^2)^2} > 0 \, , \tag{82}$$

  so $w(w')$ is strictly increasing on $[-c, c]$ with $w(-c) = -c$ and $w(c) = c$, so $w(w') \in (-c, c)$ for $w' \in (-c, c)$.[13]
- For a massless particle with $w' = \pm c$, $w = \pm c$ as well, as must be the case since light travels at speed c in all frames.

[13] You can also prove this from the properties of the tanh function by working in terms of rapidities.

Example: Multiple frames

In a frame S' an object moving with velocity w' in the x-direction is measured to have length ℓ'. The frame S' is itself moving with velocity v in the x-direction relative to another frame S. What is the object's length in S?

We cannot apply the length contraction formula Eq. (72) directly, because that involves the object's proper length. Therefore, we first calculate the object's proper length

$$\ell_0 = \gamma_{w'} \ell' \ . \tag{83}$$

In the frame S the object is moving with velocity w given by

$$w = \frac{v + w'}{1 + \frac{vw'}{c^2}} \ , \tag{84}$$

and hence has length

$$\ell = \frac{\ell_0}{\gamma_{w'}} = \frac{\gamma_w}{\gamma_{w'}} \ell' \ . \tag{85}$$

The resulting expression isn't particularly insightful, but note that if $v = -w'$ the object is stationary in S with $\gamma_{w'} = 1$ so $\ell = \ell_0$ as expected.

Example: Fizeau's Experiment

In the rest frame of a medium, e.g. water, light travels at speed c/n with $n > 1$ the refractive index. Suppose light is first shone through a stationary tank of water, and then through a tank of water moving at speed v in the direction of the light's propagation. What is difference in the light's speed measured in the lab frame in these two cases in the limit $v/c \ll 1$?

In the first case, the light's speed in the lab frame is just c/n. In the second case, its speed in the lab frame is $(v + c/n)/(1 + v/(nc))$. Hence the difference in speeds in the two cases is

$$\begin{aligned} \frac{v + \frac{c}{n}}{1 + \frac{v}{nc}} - \frac{v}{n} &\simeq \left(v + \frac{c}{n}\right)\left(1 - \frac{v}{nc}\right) - \frac{c}{n} \\ &\simeq v\left(1 - \frac{1}{n^2}\right) \ , \end{aligned} \tag{86}$$

so the light travels slightly faster in water moving in the same direction as it.

Historically, an experiment of this sort was important for ruling out the possibility that an ether existed and was dragged along by moving matter. If this was true, a much larger shift of order v, rather than $v(1 - 1/n^2)$ as predicted by Eq. (86), would have been observed. Like the Michelson–Morley experiment, this made use of an interferometer.

We can also derive the transformation of the velocity of an object moving in a general direction,

not necessarily parallel to the relative motion of two frames (which we take to still be in the x-direction).

Velocity addition formula

Consider a frame S' moving with velocity v in the x-direction relative to a frame S. Suppose an object has velocity $\mathbf{w}' = (w'_x, w'_y, w'_z)$ measured in the frame S'. Using Eq. (80) and that for a boost in the x-direction $dy' = dy$ and $dz' = dz$, the components of the object's velocity in S, $\mathbf{w} = (w_x, w_y, w_z)$, are

$$\begin{aligned} w_x &= \frac{dx}{dt} = \frac{v + w'_x}{1 + vw'_x/c^2} \\ w_y &= \frac{dy}{dt} = \frac{w'_y}{\gamma_v(1 + vw'_x/c^2)} \\ w_z &= \frac{dz}{dt} = \frac{w'_z}{\gamma_v(1 + vw'_x/c^2)} \ . \end{aligned} \tag{87}$$

To go in the opposite direction and find an object's velocity in S' given its velocity in S we just use that the velocity of S relative to S' is $-v$ in the x-direction. With this, and roles of the primed and unprimed coordinates in Eq. (87) interchanged, we obtain

$$\begin{aligned} w'_x &= \frac{-v + w_x}{1 - vw_x/c^2} \\ w'_y &= \frac{w_y}{\gamma_v(1 - vw_x/c^2)} \\ w'_z &= \frac{w_z}{\gamma_v(1 - vw_x/c^2)} \ . \end{aligned} \tag{88}$$

Example: Perpendicular velocity addition

An observer on a spaceship measures an object to have velocity $0.1c$ in the y-direction. The spaceship is moving relative to the inertial frame of an observer on Earth at velocity $0.8c$ in the x-direction. What is the velocity $\mathbf{w}$ of the object in the Earth's reference frame?

Using the formula above, with $v = 0.8c$, so $\gamma_v = 1.67$, and $\mathbf{w}' = (0, 0.1c, 0)$

$$\mathbf{w} = (0.8c, 0.06c, 0) \ . \tag{89}$$

14 "Paradoxes"

Various thought-experiments lead to so-called "paradoxes" in special relativity, i.e. results that at first sight don't seem to make sense. I stress that these do not reflect an inconsistency of the

theory, and instead simply show that our everyday intuition can fail when applied to systems with speeds close to the speed of light. I go through a few common examples in Appendix B. Some of these paradoxes are resolved by the already emphasized fact that spatially separated events that are simultaneous in one frame are not simultaneous in a boosted frame. Others (e.g. the twin paradox, which is analysed on tutorial sheet 1) are resolved by realising that if an observer can tell if they have changed inertial reference frame, since they can detect acceleration.

15 4-vectors

In preparation for what follows, let's summarise some things we already know in slightly different language. In your course on linear algebra ("vectors and matrices") you'll have seen an abstract definition of a vector. We define some particular types of vector as follows.

A ***3-vector*** $\mathbf{v}$ is a three-component vector that transforms in the same way as space coordinates under spatial rotations: $\mathbf{v}' = R\mathbf{v}$ where R is the rotation matrix. For example, the usual velocity of a particle $(dx/dt, dy/dt, dz/dt)$ is a 3-vector, known as the ***3-velocity***, as is the acceleration, called the ***3-acceleration***. The dot product of any two 3-vectors is invariant under rotations

$$\mathbf{w}'.\mathbf{v}' = \mathbf{w}R^T R\mathbf{v} = \mathbf{w}.\mathbf{v} \ . \tag{90}$$

Returning to including time, we're used to packaging the ct and x coordinates of an event into a single vector (ct, x). To be general we include y and z to make a vector with four elements. Usually we denote such a vector with a capital letter, e.g.

$$X = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \ . \tag{91}$$

Although we proved it only for boosts in the x-direction, the transformation of the coordinates from one inertial frame S to another S' is always a linear transformation, so it can be written

$$X' = \Lambda X \ , \tag{92}$$

where Λ is a 4×4 matrix corresponding to the specific transformation, generalising Eq. (49).

The linearity of Lorentz transformations follows from the assumption that space and time are homogeneous, which implies that translating an object so its ends move from $x = 0$ and $x = \ell$ to $x = c$ and $x = c + \ell$ cannot change its length measured in a frame S'. This shows that $\partial x'/\partial x$ must be a constant. Similar arguments can be made for the time coordinates.

The normal squared length of X, $c^2t^2 + x^2 + y^2 + z^2$, is not Lorentz invariant (you can easily check this for a boost). Therefore, instead of using the standard dot product, we define $X \cdot X$ to be the invariant interval squared. This is known as the ***Minkowski inner product***. In matrix form, it can be written

$$X \cdot X := X^T \eta X = c^2t^2 - x^2 - y^2 - z^2 \ , \tag{93}$$

where

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \ . \tag{94}$$

We derived the Lorentz transformations from the requirement that the interval is invariant for all X (in fact, this can serve as the *definition* of Lorentz transformations). Therefore,

$$\begin{aligned} X' \cdot X' &\equiv X'^T \eta X' \\ &= X^T \Lambda^T \eta \Lambda X \ , \end{aligned} \tag{95}$$

so the matrices Λ satisfy

$$\Lambda^T \eta \Lambda = \eta \ , \tag{96}$$

(c.f. rotations, for which $R^T I R = I$ where I is the identity matrix).

Now something new:

Four vectors

A ***four vector*** is any vector with four components that transforms between inertial reference frames in the same way as Eq. (92).

Of course, spacetime position vectors $X = (ct, x, y, z)^T$ are themselves 4-vectors. From the way that rotations act on 4-vectors, the spatial part of a 4-vector (e.g. (x, y, z)) is a 3-vector.

Caution: Given the transformation Eq. (92)*, all 4-vectors should be written as column vectors (or as row vectors transposed). In books you might see 4-vectors written as row vectors, in which case it's implicit that before transforming them between frames they should be transposed.*

Example: Something that is not a 4-vector

Consider the worldline of a particle in a frame S', with its spatial coordinates parameterised by t'. We can construct a vector with four elements $dX'/dt' = (c, dx'/dt', dy'/dt', dz'/dt')^T$. Using the analogue of Eq. (80), in terms of the coordinates of a frame S in which S' moves with

velocity v in the x-direction, this vector is

$$\begin{pmatrix} c \\ dx'/dt' \\ dy'/dt' \\ dz'/dt' \end{pmatrix} = \begin{pmatrix} c \\ \frac{-v+dx/dt}{1-\frac{v}{c^2}dx/dt} \\ \frac{1}{\gamma_v\left(1-\frac{v}{c^2}dx/dt\right)}dy/dt \\ \frac{1}{\gamma_v\left(1-\frac{v}{c^2}dx/dt\right)}dz/dt \end{pmatrix} \neq \Lambda \begin{pmatrix} c \\ dx/dt \\ dy/dt \\ dz/dt \end{pmatrix} , \tag{97}$$

where Λ is the Lorentz transformation from S to S'. Therefore, the dX/dt is not a 4-vector. (Once we have shown (below) that $X \cdot X$ is Lorentz invariant for any 4-vector, we can come to the same conclusion by noting that $dX/dt \cdot dX/dt = c^2 - |\mathbf{u}|^2$ is not invariant.)

Example: Something that is a 4-vector

The trouble in the previous example stemmed from the transformation law of "dt'". This suggests a way to obtain a 4-vector. Consider

$$\frac{dX'}{d\tau} = \begin{pmatrix} c\,dt'/d\tau \\ dx'/d\tau \\ dy'/d\tau \\ dz'/d\tau \end{pmatrix} , \tag{98}$$

where τ is the proper time, with $d\tau$ given by Eq. (77). Since τ is Lorentz invariant

$$\begin{pmatrix} c\,dt'/d\tau \\ dx'/d\tau \\ dy'/d\tau \\ dz'/d\tau \end{pmatrix} = \begin{pmatrix} c\ dt'/dt\ dt/d\tau \\ dx'/dx\ dx/d\tau \\ dy'/dy\ dy/d\tau \\ dz'/dz\ dz/d\tau \end{pmatrix} = \Lambda \begin{pmatrix} c\,dt/d\tau \\ dx/d\tau \\ dy/d\tau \\ dz/d\tau \end{pmatrix} . \tag{99}$$

where we used that dt'/dt etc. gives the elements of Λ. Evidently, $dX/d\tau$ is a 4-vector.

Another way to see that $dX/d\tau$ is 4-vector is that $\delta X = (\delta t, \delta x, \delta y, \delta z)^T$ is a 4-vector, and dividing by the invariant $\delta\tau$ (and taking the limit $\delta\tau \to 0$) does not affect how it transforms.

We define the scalar product of any 4-vector, Y say, as in Eq. (93). We immediately see that $Y \cdot Y$ is Lorentz invariant, because its value in a frame S'

$$\begin{aligned} Y' \cdot Y' &\equiv Y^T \Lambda^T \eta \Lambda Y \\ &= Y^T \eta \Lambda Y \\ &= Y \cdot Y \ . \end{aligned} \tag{100}$$

A more general version of the same calculation leads to the crucial result:

Invariant scalar product

The scalar product of any two 4-vectors, X and Y, defined by

$$X \cdot Y \equiv X^T \eta Y \ , \tag{101}$$

is Lorentz invariant: $X' \cdot Y' = X^T \Lambda^T \eta \Lambda Y = X^T \eta Y = X \cdot Y$.

Writing the components of 4-vectors as $X = (X^0, X^1, X^2, X^3)^T$ etc., we have

$$X \cdot Y = X^0 Y^0 - X^1 Y^1 - X^2 Y^2 - X^3 Y^3 \ . \tag{102}$$

(For this course you don't need to know why the indices are written in the up position, but see the red comment below).

Note that we start counting the components of a 4-vector from 0, so the parts that correspond to space for a coordinate vector are labelled $1 \to 3$. Analogously, we refer to the component 0 of a general 4-vector as its time component and the $1 \to 3$ as its spatial components.

We use the terminology that the *square* of a 4-vector, X say, is $X^2 = X \cdot X$.

It is easy to check that the scalar product is commutative $X \cdot Y = Y \cdot X$, is distributive $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$, and for any scalar c satisfies $(cX) \cdot Y = c(X \cdot Y)$. Therefore, for example, $(X + Y) \cdot (X + Y) = X \cdot X + Y \cdot Y + 2X \cdot Y$.

As we saw above, one way of producing a new 4-vector from a given 4-vector is to differentiate it with respect to an invariant quantity (the proof is as in Eq. (98)). The sum of two 4-vectors is automatically a 4-vector, as is a 4-vector multiplied or divided by a Lorentz invariant quantity.

Just like for the spacetime coordinate 4-vectors discussed near Eq. (57), a general 4-vector Y is classified as time-like, space-like or light-like depending on whether $Y \cdot Y$ is greater than 0, less than 0 or equal to 0. The results in Section 9 about the existence of reference frames in which position 4-vectors take specific forms hold also for other types of 4-vector. For example, for any time-like 4-vector there always exists a frame in which its space-components vanish.

In books and courses in later years, you'll come across a useful notation, called index notation. In this we write $X \cdot Y = \eta_{\mu\nu} X^\mu Y^\nu$, with $\eta_{\mu\nu}$ denoting the components of the matrix η and X^μ and Y^ν the components of X and Y. A summation over repeating indices is implied.

We then *define* the index down version of e.g. Y by $Y_\mu = \eta_{\mu\nu} Y^\nu$. Thinking of (Y_μ) as a row vector $(Y_\mu) = (Y^0, -Y^1, -Y^2, -Y^3)$, under a Lorentz transformation $(Y'_\mu) = (Y_\mu)\Lambda^{-1}$ (to prove this, use that $(Y_\mu)(X^\mu)$ is invariant).

The usual summation convention you'll have met in other courses is modified such that only indices that appear once up and once down are summed over, so $Y \cdot X = Y_\mu X^\mu = X_\mu Y^\mu$. This

guarantees that expressions constructed from 4-vectors are Lorentz invariant (if all indices are contracted), or transform as some non-trivial representation of the Lorentz group (if some indices remain uncontracted).

Using the chain rule you can show that

$$(\partial_\mu) := \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) , \tag{103}$$

transforms like an index down vector: $(\partial'_\mu) = (\partial_\mu)\Lambda^{-1}$. Consequently, the index up version

$$(\partial^\mu) = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right)^T , \tag{104}$$

transforms like a 4-vector and $\partial_\mu\partial^\mu = \frac{1}{c^2}\partial_t^2 - \partial_x^2 - \partial_y^2 - \partial_z^2$ is Lorentz invariant. This is a slick way of seeing that the wave equation, Eq. (18), is Lorentz invariant.

For 4-vectors there is not an operation analogous to the 3-vector cross product. Instead what can be defined is an object $\epsilon_{\mu\nu\rho\sigma}X^\rho Y^\sigma$, which has two indices and is anti-symmetric ($\epsilon_{\mu\nu\rho\sigma}$ is the 4-dimensional Levi-Civita tensor).

16 4-velocity and 4-acceleration

The example of a 4-vector introduced in Eq. (98) is sufficiently important that we give it a name.

The 4-velocity

Consider a time-like worldline with coordinates $X(\tau)$ parameterised by the proper time τ along it. Let E be an event on the worldline. We define the ***4-velocity*** at E to be

$$\begin{aligned} U &= \frac{dX}{d\tau} \\ &= \left(c\frac{dt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau}\right)^T \\ &= (\gamma_u c, \gamma_u \mathbf{u})^T , \end{aligned} \tag{105}$$

where all derivatives are evaluated at E, $\mathbf{u} = (dx/dt, dy/dt, dz/dt)$ is the 3-velocity, $\gamma_u = (1 - |\mathbf{u}|^2/c^2)^{-1/2}$ is the Lorentz γ factor, and we used that e.g. $dx/d\tau = dx/dt\ dt/d\tau$ and $dt/d\tau = \gamma$.

The 4-velocity is a 4-vector, so we can transform it between frames using Eq. (92).

Exercise: use the transformation law of the 4-velocity between frames to re-derive the velocity addition formula, Eq. (81).

Because $U \cdot U$ is a Lorentz invariant, we can find its value using the following trick (which often saves effort).

Useful trick: evaluating an invariant in a convenient frame

Suppose we want to know the value of a Lorentz invariant, e.g. $U \cdot U$. Often, we can do so by calculating its value in a frame where $U' = \Lambda U$ takes a particularly simple form enabling $U \cdot U = U' \cdot U'$ to be easily calculated.

More generally, suppose we want to know the value of some physical quantity. If we can write this in terms of Lorentz invariant quantities, e.g. $X \cdot Y$ where X and Y are 4-vectors, then these can be evaluated in whatever frame is most convenient. Typically this is either a frame where the 4-vectors take a simple form, or the frame in which we already have information.

Example: 4-velocity squared

Find $U \cdot U$ for any 4-velocity U.

To do so, we go to the frame S' in which a particle travelling along the worldline is instantaneously at rest at time t, so $U' = (c, 0, 0, 0)^T$. We therefore have $U \cdot U = U' \cdot U' = c^2$. This can also be checked from the general expression Eq. (105).

Note: since e.g. $X \cdot Y = X' \cdot Y'$, I will often write that "we evaluate $X \cdot Y$ in a frame S'". This is short-hand for "we calculate $X' \cdot Y'$, where X', Y' are the 4-vectors transformed to the frame S', and use the fact that $X' \cdot Y' = X \cdot Y$.

Caution: Although the sum of two 4-velocities, U_1 and U_2 say, is a 4-vector it is not a 4-velocity and this is certainly not how addition of velocities works. To see this note that $(U_1 + U_2)^2 = 2c^2 + 2U_1 \cdot U_2 \geq 2c^2$ *(exercise: prove this)*, whereas any 4-velocity satisfies $U \cdot U = c^2$.

Differentiating the 4-velocity with respect to proper time again we obtain another 4-vector.

The 4-acceleration

The ***4-acceleration*** A is the derivative of the 4-velocity with respect to proper time

$$A = \frac{dU}{d\tau} \ . \tag{106}$$

Using that $dt/d\tau = \gamma$ we have

$$A = \gamma \frac{d}{dt}\left(\gamma \frac{dX}{dt}\right) \ . \tag{107}$$

Eq. (107) can be simplified further using

$$\frac{d\gamma_u}{dt} = \frac{d\gamma_u}{du}\frac{du}{dt} = \frac{\gamma^3 u}{c^2}\frac{du}{dt} \ , \tag{108}$$

and

$$u\frac{du}{dt} = \frac{1}{2}\frac{d}{dt}\left(u^2\right) = \frac{1}{2}\frac{d}{dt}\left(\mathbf{u}.\mathbf{u}\right) = \mathbf{u}.\mathbf{a} \ . \tag{109}$$

where $u = |\mathbf{u}|$. Expanding the derivative in Eq. (107), we obtain

$$\begin{aligned} A &= \gamma^4 \frac{u}{c^2}\frac{du}{dt}\left(c, \mathbf{u}\right)^T + \gamma^2 \left(0, \mathbf{a}\right)^T \\ &= \gamma^2 \left(\frac{\mathbf{u}.\mathbf{a}}{c}\gamma^2, \frac{\mathbf{u}.\mathbf{a}}{c^2}\gamma^2\mathbf{u} + \mathbf{a}\right)^T \ , \end{aligned} \tag{110}$$

where as above $\mathbf{u}$ is the 3-velocity, and $\mathbf{a} = d\mathbf{u}/dt$ is the 3-acceleration. In books you will find various equivalent expressions for A.

In the instantaneous rest frame of a particle, $\gamma = 1$ and $\mathbf{u} = \mathbf{0}$, so $A = (0, \mathbf{a}_{\rm inst})^T$, where $\mathbf{a}_{\rm inst}$ denotes the acceleration measured in this frame. Hence the invariant $A \cdot A$ is given by

$$A \cdot A = -|\mathbf{a}_{\rm inst}|^2 \ . \tag{111}$$

Moreover in the instantaneous rest frame $U = (c, 0, 0, 0)^T$, so $U \cdot A = 0$ (i.e. the 4-velocity is always orthogonal to the 4-acceleration). The latter fact can also be deduced by differentiating $U \cdot U = c^2$ with respect to proper time.

Suppose that we are in a frame in which the 3-velocity and 3-acceleration are perpendicular, i.e. $\mathbf{u}.\mathbf{a} = 0$. Then, using Eq. (110),

$$A \cdot A = -\gamma^4 |\mathbf{a}|^2 \ . \tag{112}$$

Hence, the acceleration in this frame is related to that in the particle's instantaneous rest frame by

$$|\mathbf{a}| = \gamma^{-2} |\mathbf{a}_{\rm inst}| \, . \tag{113}$$

A similar calculation can be done in the case that the 3-acceleration is parallel to the 3-velocity, i.e. $\mathbf{u}.\mathbf{a} = \pm |\mathbf{u}| \, |\mathbf{a}|$. Again using Eq. (110), a few lines of calculation shows that

$$\begin{aligned} A \cdot A &= -\gamma^6 |\mathbf{a}|^2 \\ \Longrightarrow \ |\mathbf{a}| &= \gamma^{-3} |\mathbf{a}_{\rm inst}| \, . \end{aligned} \tag{114}$$

Example: 4-velocities and 4-accelerations

The worldline of a particle in a frame S is

$$\begin{aligned} x(t) &= a\sin(\omega t) \\ y(t) &= a\cos(\omega t) \\ z(t) &= 0 \ . \end{aligned} \tag{115}$$

Calculate the particle's 4-velocity and 4-acceleration.

In this frame, the particle's 3-velocity is $\mathbf{u}(t) = (a\omega\cos(\omega t), -a\omega\sin(\omega t), 0)$ and

$$\gamma(t) = \left(1 - a^2\omega^2/c^2\right)^{-1/2} \ . \tag{116}$$

From Eq. (105), its 4-velocity is therefore

$$U = \frac{1}{\sqrt{1 - a^2\omega^2/c^2}} \left(c, a\omega\cos(\omega t), -a\omega\sin(\omega t), 0\right)^T \ . \tag{117}$$

The 4-acceleration can be found either by differentiating Eq. (117) or from Eq. (110). Either way, since $du/dt = 0$ we obtain

$$A = \frac{1}{1 - a^2\omega^2/c^2} \left(0, -a\omega^2\sin(\omega t), -a\omega^2\cos(\omega t), 0\right)^T \ . \tag{118}$$

As expected $U \cdot A = 0$.

Example: Constant acceleration

Consider a particle moving along the x-axis with constant acceleration in its instantaneous rest frame of a_{inst} in the x-direction. Determine the trajectory of the particle in an inertial frame S in which the particle is stationary at $x = 0$ at $t = 0$.

There are different ways to solve this problem. Perhaps the most straightforward is via Eq. (114), which shows that

$$\frac{dv}{dt} = \left(1 - v^2/c^2\right)^{3/2} a_{\text{inst}} \ . \tag{119}$$

Integrating

$$\begin{aligned} \int_0^{v(t)} \frac{dv}{(1-v^2/c^2)^{3/2}} &= \int_0^t a_{\rm inst} \\ \implies \frac{v}{\left(1-\frac{v^2}{c^2}\right)^{1/2}} &= a_{\rm inst} t \\ \implies v = \frac{a_{\rm inst} t}{\sqrt{1+a_{\rm inst}^2 t^2/c^2}} &\ . \end{aligned} \tag{120}$$

Integrating again and fixing the constant of integration

$$x = \frac{c^2}{a_{\rm inst}} \sqrt{1+\frac{a_{\rm inst}^2 t^2}{c^2}} - \frac{c^2}{a_{\rm inst}} \ . \tag{121}$$

This is a hyperbolic trajectory in a spacetime diagram.

Plugging the expression for v in Eq. (120) into γ, we find

$$\gamma(t) = \sqrt{1+a_{\rm inst}^2 t^2/c^2} \ . \tag{122}$$

Hence the relation between the particle's proper time and the time elapsed in the reference frame in which it is stationary at $t=0$ is

$$\begin{aligned} \tau &= \int_0^t \frac{dt}{\sqrt{1+a_{\rm inst}^2 t^2/c^2}} \\ &= \frac{c}{a_{\rm inst}} {\rm arcsinh}\left(t a_{\rm inst}/c\right) \ . \end{aligned} \tag{123}$$

i.e

$$t = \frac{c}{a_{\rm inst}} \sinh\left(\tau a_{\rm inst}/c\right) \ , \tag{124}$$

so

$$x = \frac{c^2}{a_{\rm inst}} \left(\cosh\left(\tau a_{\rm inst}/c\right) - 1\right) \ . \tag{125}$$

Note that the 3-acceleration in the frame S decreases with time, such that the particle's speed remains smaller than the speed of light. In more detail, from Eq. (121) for small $a_{\rm inst}t/c$ we have that $x \simeq \frac{1}{2} a_{\rm inst} t^2$ as expected in Newtonian mechanics, while at large $a_{\rm inst}t/c$ we have $x \simeq ct$, i.e. the particle's speed in S tends to the speed of light.

A more dramatic fact is that the growth in x and t are exponential in proper time τ, as can be seen by plugging Eq. (124) into Eq. (121). For example, if the acceleration is $10\,{\rm ms}^{-2}$ we have that as τ goes from 0 to 10 years, t goes from 0 to $\frac{1}{2}e^{10}$ years (about 11000 years). *Exercise: check this.* If a rocket was able to maintain g-acceleration for 10 years, astronauts could travel of order 10^4 years into the future relative to the Earth's reference frame.

Comment: One sometimes reads that special relativity can only deal with uniform motion, and that general relativity is needed to treat acceleration. These statements are misleading. The real restriction in special relativity is that the coordinate systems we use are inertial coordinate systems.[14] *But, as we have seen in the examples above, this does not prevent us from discussing accelerating particles. Indeed, Einstein's 1905 paper was titled 'On the Electrodynamics of Moving Bodies', and dynamics is all about force and acceleration.*

17 Relativistic energy and momentum

The modern approach to deriving the relativistic analogues of Newton's 2nd law and the conserved energy and momenta is to start with an action and apply the Euler-Lagrange equations and Noether's theorem. Unfortunately, you haven't yet met this formalism in Newtonian mechanics, so we can't take this route for relativistic dynamics! Instead we will *postulate* plausible relativistic laws that satisfy the principle of relativity under Lorentz transformations and reduce to Newtonian mechanics in the limit $v/c \ll 1$. Ultimately, these are validated by comparison with experiment. Still, it's worth keeping in the back of your mind that the things we postulate can be derived from a "deeper" principle.

Consider some physical object such as a particle (usually I'll just call everything a particle but these don't need to be elementary particles and all the results apply also to complicated extended objects). We define the particle's mass m to be its mass as measured in its instantaneous rest frame (this definition is Lorentz invariant by construction). Sometimes m is called the particle's *rest mass* (see the comment below Eq. (130)).

Caution: Standard terminology is to call any particle that has a non-zero mass "massive". This does not imply that the mass is large.

4-momentum

We define the ***4-momentum*** of a massive particle by

$$\begin{aligned} P &= mU \\ &= m\gamma_u \left(c, \mathbf{u}\right)^T \ , \end{aligned} \tag{126}$$

where U is its 4-velocity and $\mathbf{u}$ is its 3-velocity. P is a 4-vector.

[14] We must also neglect the effect of gravity. It turns out that the key idea for incorporating gravity is exploiting the freedom to use any coordinates we like, and this is the origin of the word 'general' in general relativity.

To see that the 4-momentum of a particle P is a interesting quantity, consider its Taylor expansion in the limit $u/c \ll 1$, e.g. with $\mathbf{u}$ aligned in the x-direction. Then $\gamma_u \simeq 1 + \frac{1}{2}u^2/c^2$ and

$$P \simeq \begin{pmatrix} mc + \frac{1}{2}mu^2/c \\ mu + \frac{1}{2}mu^3/c^2 \\ 0 \\ 0 \end{pmatrix} . \tag{127}$$

The spatial part of P is the Newtonian momentum up to a correction that vanishes in the limit $u/c \to 0$. We therefore interpret this as the relativistic version of momentum, which we call the ***3-momentum*** $\mathbf{p} = m\gamma\mathbf{u}$. This differs from the expression for momentum in Newtonian mechanics by the factor of γ.

What about the time component of P? The first term in P^0 in Eq. (127) is just a constant, while the second $(\frac{1}{2}mu^2/c)$ is the Newtonian kinetic energy divided by c. We therefore make a jump and postulate that the time component of P is the relativistic energy of a massive particle E divided by c. Consequently:

Relativistic energy and momentum

The energy E and 3-momentum $\mathbf{p}$ of a particle are given by the components of the 4-momentum P

$$P \equiv (E/c, \mathbf{p})^T . \tag{128}$$

Since $P = mU = (m\gamma c, m\gamma\mathbf{u})^T$, we have

$$E = \gamma mc^2 , \qquad \text{and} \qquad \mathbf{p} = m\gamma\mathbf{u} . \tag{129}$$

It can also be useful to know that, from Eq. (129),

$$\frac{|\mathbf{p}|}{E} = \frac{|\mathbf{u}|}{c^2} . \tag{130}$$

The energy of a stationary particle is $E = mc^2$, which we interpret as a contribution to a particle's energy from its rest mass. Conversely, the ***kinetic energy*** of a particle is $E - mc^2 = (\gamma - 1)mc^2$.

In the limit $u/c \to 1$ the energy and 3-momentum of a particle both tend to infinity because $\gamma_u \to \infty$. This justifies our claim back in Section 5 that the speed of massive particles is always strictly less than the speed of light: it would take infinite energy to accelerate a massive particle to the speed of light.

Caution: Sometimes people treat the combination γm as a velocity-dependent relativistic mass. With this interpretation, the relativistic mass of a particle diverges as it approaches the speed of

light. Although perfectly consistent (the maths is unchanged), this way of thinking usually does not add any insights. We therefore stick with our approach of only using the rest mass m and keeping factors of γ explicit.

It is often useful to have an expression for the energy of a particle involving the particle's momentum rather than its speed. This can be obtained by solving $p = m\gamma_u u$ for u and substituting into $E = m\gamma c^2$. A much nicer way to do the same thing is by using the Lorentz invariant $P \cdot P$.

Energy momentum relation

In the rest frame of a particle of mass $m > 0$ we have $P = (mc, 0, 0, 0)$, hence $P \cdot P = m^2c^2$. Since $P \cdot P$ is Lorentz invariant, in a general frame

$$P \cdot P = E^2/c^2 - |\mathbf{p}|^2 = m^2c^2 \tag{131}$$

i.e.

$$E^2 = m^2c^4 + |\mathbf{p}|^2c^2 \ . \tag{132}$$

This is the complete version of the famous $E = mc^2$ (which applies to stationary particles).

Since $m > 0$, P is a time-like 4-vector with positive time-like component. Hence, there exists an inertial frame in which, instantaneously, $\mathbf{p} = \mathbf{0}$.

For a system of several particles the total 4-momentum of the system is the sum of the 4-momenta of the individual particles $P_{\text{tot}} = \sum_i P_i$. The time component of P_{tot} is interpreted as the total energy of the system and the spatial components the total momentum. Since the total 4-momentum is a 4-vector, its values in different frames are related by the Lorentz transformation.

We now make a crucial postulate.

Conservation of 4-momentum

In the absence of external forces, the total 4-momentum of a system of particles, P_{tot}, is conserved in all frames (i.e. it is independent of time $dP_{\text{tot}}/dt = 0$).

Conservation of total 4-momentum means that each of the individual components of P_{tot} is conserved, so the total energy and each component of the total 3-momentum are conserved.

The time-like (energy) and space-like (momentum) components of the total 4-momentum are the conserved charges that, via Noether's theorem, result from invariance under translations in time and space.

In Newtonian mechanics, we have conservation of energy, momentum, and total mass. Crucially, in special relativity we no longer have conservation of total rest mass, and instead we only have conservation of total energy. (One could *postulate* that total rest mass is also conserved in

special relativity but this is found to be experimentally false and unlike 4-momentum conservation, which arises from Noether's theorem, there is no theoretical reason it should be the case.)

This is why the contribution to a particle's energy from its rest mass is crucial. If the total rest mass was conserved it would just be an irrelevant shift by a constant. However, since only total energy is conserved, energy in the form of mass can be converted into other forms. Nuclear reactions provide a particularly dramatic example of this, and we will see its importance for particle physics in the next Section.

Example: Time dilation II

The muon is an unstable subatomic particle with a mass of 106 MeV/c^2 and a mean lifetime in its rest frame of $t_\mu = 2.2 \times 10^{-6}$ s. A possible option for a future particle collider involves colliding high energy muons. What is the lab-frame lifetime of a muon beam of energy 5 TeV? How many laps of a $l = 30$ km ring would a bunch of such muons complete before 50% have decayed?

The γ factor of such muons is

$$\gamma = E/(mc^2) \ , \tag{133}$$

corresponding to a speed

$$u/c = \sqrt{1 - m^2c^4/E^2} \ . \tag{134}$$

Accounting for time dilation, the proper time for the muon beam to orbit the ring is

$$\tau_{\rm lap} = \frac{l}{\gamma u} \ . \tag{135}$$

Hence the number of laps N before 50% have decayed satisfies

$$\begin{aligned} 0.5 &= \exp\left(-N\tau_{\rm lap}/t_\mu\right) \\ \implies N &= \frac{t_\mu \gamma u}{l} \log(2) \\ &= 720 \ . \end{aligned} \tag{136}$$

Since $u/c \simeq 1 + \mathcal{O}(m^2/E^2)$, when plugging numbers into the middle line of Eq. (136) we can take the explicit factor of u to be c to the accuracy we are working, saving a bit of effort. Be careful when making such approximations though.

Side-note: You've probably seen this somewhere already, but if not it's worth checking that if the mean lifetime of a particle is t_μ then the fraction of particles surviving a time t is e^{-t/t_μ}. We used this in Eq. (136)*.*

Example: Energy and momenta in different frames

In the lab frame S a particle of mass m has velocity $u = 0.9c$ in the x-direction. Consider a second frame S' that is moving relative to S with velocity $v = 0.7c$ in the x-direction. What is the energy and momentum of the particle in S and S'?

In the lab frame, the particle's energy is $E = \gamma_u mc^2 = 2.3mc^2$, and its momentum is $|\mathbf{p}| = 2.1mc$ in the x-direction. We can find the energy and momentum in S' using the Lorentz transformation

$$\begin{pmatrix} E'/c \\ p'_x \end{pmatrix} = \begin{pmatrix} \gamma_v & \frac{-v}{c}\gamma_v \\ \frac{-v}{c}\gamma_v & \gamma_v \end{pmatrix} \begin{pmatrix} E/c \\ p_x \end{pmatrix} . \tag{137}$$

Hence

$$\begin{aligned} E' &= \gamma_v E - \gamma_v v p \\ &= 1.2mc^2 , \end{aligned} \tag{138}$$

and

$$\begin{aligned} p'_x &= -\frac{v}{c^2}\gamma_v E + \gamma_v p_x \\ &= 0.64mc . \end{aligned} \tag{139}$$

We can check that the results in S and S' both satisfy $E^2 = p^2c^2 + m^2c^4$ to within rounding.

Exercise: write a general expression for E', and plot this as a function of u for a few different values of v, e.g. using Mathematica. Check that your results make sense physically.

Example: Invariants

A particle with 4-momentum P is detected by an observer whose 4-velocity is V. Working in the rest frame of the observer, express $P \cdot V$ in terms of the mass of the particle and its speed in the observer's rest frame u. Show that

$$\frac{u}{c} = \sqrt{1 - \frac{(P \cdot P)\, c^2}{(P \cdot U)^2}} . \tag{140}$$

In the observers rest frame their own 4-velocity $V = (c, 0, 0, 0)^T$ so $P \cdot V = E'$, where E' is the particle's energy in this frame. Hence, the invariant

$$P \cdot V = \gamma_u mc^2 = \frac{mc^2}{\sqrt{1 - u^2/c^2}} . \tag{141}$$

Solving Eq. (141) for u we have

$$\begin{aligned} \frac{u}{c} &= \sqrt{1 - \frac{m^2c^4}{(P \cdot V)^2}} \\ &= \sqrt{1 - \frac{(P \cdot P)c^2}{(P \cdot V)^2}} , \end{aligned} \tag{142}$$

where in the last line we used that $P \cdot P = m^2c^2$. Since it involves only invariants, the expression on the right-hand side can be evaluated in any convenient frame.

Finally, note that $E^2 = p^2c^2 + m^2c^4$ is valid only for free particles, because it does not include contributions from potential energies. For example, in the presence of electromagnetic fields the energy of a charged particle involves an additional term. When we subsequently use 4-momentum conservation to study collisions of particles, what we really do is compare the total 4-momenta long before and long after a collision when all the contributions from potential energies are negligible.

You might read about tachyons, which are hypothetical particles with imaginary mass, such that $m^2 < 0$. These would have $P \cdot P < 0$, which from Eq. (128) means they travel at speeds $u > c$ (i.e. on space-like trajectories), enabling faster-than-light communication. Although they are consistent within special relativistic particle mechanics, tachyons are not consistent with quantum field theory. In quantum field theory, particles are excitations of fields that permeate spacetime. A negative mass squared indicates that we are doing a Taylor expansion of the field around the wrong point; i.e. it corresponds to an instability of the vacuum and we could lower the energy of the system by creating more and more particles. Instead, we should expand around a different field value. In the 4th year QFT option, you will learn that the Higgs mechanism is an example of this phenomenon.

18 Massless particles and Doppler shift

What is the 4-momentum P_γ of a photon? Recall from Eq. (78) that proper time, and therefore 4-velocity $U = dX/d\tau$, is only defined for time-like trajectories. However, photons travel on light-like trajectories. Therefore, we cannot define P_γ using Eq. (126) and a different approach is needed.

It's reasonable to seek a definition that generalises Eq. (131) to the $m = 0$ case, i.e.

$$P_\gamma \cdot P_\gamma = \frac{E^2}{c^2} - |\mathbf{p}|^2 = 0 \ , \tag{143}$$

so that a photon's 4-momentum is light-like. This means the photon's energy and momentum are related by $E = |\mathbf{p}|c$, so Eq. (130) holds also with $|\mathbf{u}| = c$.

From quantum mechanics we know that the energy of a photon is related to its angular frequency $\omega = 2\pi c/\lambda$ by

$$E = \hbar\omega = \frac{2\pi\hbar c}{\lambda} \ , \tag{144}$$

where λ is its wavelength. We therefore take the 0th component of P to be $\hbar\omega$, leading to:

4-momentum of massless particles

For massless particles

$$P = \frac{\hbar\omega}{c}\,(1, \hat{\mathbf{n}})^T \ , \tag{145}$$

where $\hat{\mathbf{n}}$ is a unit vector in the direction of the particle's propagation.

In terms of the photon frequency $\nu = \omega/(2\pi)$ rather than its angular frequency

$$P = \frac{h\nu}{c}(1, \hat{\mathbf{n}})^T \ . \tag{146}$$

Experiments confirm that, with this definition of photon 4-momentum, conservation of energy and momentum indeed holds for systems involving photons .

Eq. (143) also clarifies the status of other massless particles (the only known example of which is the graviton). We postulate that Eq. (143) and $|\mathbf{p}|/E = |\mathbf{u}|/c^2$ hold for any such particle, so they necessarily travel at the speed of light with $|\mathbf{u}| = c$.[15]

As yet, you might question whether the 4-momentum of a massless particle defined by Eq. (145) is necessarily a 4-vector. To see that this is the case, let's go back to thinking of light as waves propagating according to the wave equation Eq. (18). One solution of this is

$$E_y = A\cos\left(\frac{\omega}{c}(ct - \hat{\mathbf{n}}.\mathbf{x})\right) \ . \tag{147}$$

where A is a constant. Eq. (147) can be written in terms of the ***frequency 4-vector***

$$K = P/\hbar = \frac{\omega}{c}\,(1, \hat{\mathbf{n}})^T \ , \tag{148}$$

and the position 4-vector $X = (ct, \mathbf{x})^T$ as

$$E_y = A\cos(K \cdot X) \ . \tag{149}$$

The phase of a wave at a given spacetime event is Lorentz invariant (e.g. if $E_y = 0$ all observers agree on this). Therefore $K \cdot X$ is Lorentz invariant, and hence K is indeed a 4-vector, and so is $P = \hbar K$.

[15] Another way to argue the same thing is that if a massless particle travelled at less than the speed of light we could go to a reference frame in which it was at rest. However, since $m^2 = 0$ we would have $P = (0, 0, 0, 0)$ in this frame and therefore P would be zero in every frame.

Doppler shift

Longitudinal Doppler shift

Suppose that in a frame S a beam of photons is travelling in the positive x-direction with frequency ω. In a frame S' moving relative to S with velocity v in the x-direction the photons have frequency

$$\omega' = \omega\sqrt{\frac{1-v/c}{1+v/c}}\,. \tag{150}$$

Proof: The Lorentz transformation of the photon's 4-momentum from S to S' is (cancelling a common factor of $\hbar/c$)

$$\begin{pmatrix}\omega'\\ \omega'\end{pmatrix} = \begin{pmatrix}\gamma_v & \frac{-v}{c}\gamma_v\\ \frac{-v}{c}\gamma_v & \gamma_v\end{pmatrix}\begin{pmatrix}\omega\\ \omega\end{pmatrix}\,, \tag{151}$$

hence,

$$\omega' = \gamma_v\left(1-\frac{v}{c}\right)\omega = \omega\sqrt{\frac{1-v/c}{1+v/c}}\,. \tag{152}$$

Note: Eq. (150) can also be derived analysing invariants, similarly to Eq. (140). To do so equate $P\cdot V$ evaluated in the two frames, where P is the photon's 4-momentum and V is the 4-velocity of the frame S'. It can also be derived from length contraction and time dilation.

For $v>0$, according to an observer in S, an observer in S' moves away from the photons. We therefore expect the observer in S' to measure the photons to have a larger wavelength and hence smaller frequency than in S (i.e. they are redshifted). This is consistent with Eq. (150), since $\omega'<\omega$ for $v>0$. Meanwhile, from the perspective of S' an observer in S is moving towards the photons, so is consistently expected to measure $\omega>\omega'$.

Conversely, for $v<0$, from the perspective of S, an observer in S' is moving towards the photons so we expect $\omega'>\omega$ (i.e. the photons are blueshifted), as predicted by Eq. (150).

In the limit $v\ll c$, Eq. (152) reduces to $\omega'/\omega\simeq 1-v/c$.

There is a more complicated expression for ω' when the relative velocity of the frames is not parallel to the photons. This is most easily derived using the approach of considering invariants mentioned above.

One can easily check that in the limit $v/c\ll 1$ Eq. (150) reduces to $\omega'/\omega = 1-v/c+\mathcal{O}(v^2/c^2)$.

19 Centre of momentum frame

Centre of momentum frame

For any system of at least two particles there exists an inertial reference frame in which the total 4-momentum of the system

$$P_{\text{tot}} = (E_{\text{COM}}/c, 0, 0, 0)^T \ , \tag{153}$$

i.e. with zero total 3-momentum. Such a frame is called the ***centre-of-momentum frame*** (COM).

Being careful, a COM frame does not exist for a system of only massless particles that are all travelling in exactly the same direction. A COM frame also exists for any single massive particle, but not for a single massless particle.

Note: the centre of momentum frame is often called the centre of mass frame (this is arguably slightly misleading terminology).

Although intuitively plausible, we should prove that the COM frame exists. We do this for a system of massive particles leaving the inclusion of massless particles as an exercise. First we derive a small lemma.

Lemma: The sum of two time-like vectors $A = (A^0, \mathbf{a})$ and $B = (B^0, \mathbf{b})$ with positive time-component is time-like, because $(A+B)^2 = A^2 + B^2 + 2A \cdot B$ with $A^2 > 0$, $B^2 > 0$ and

$$\begin{aligned} A \cdot B &= (A^0 + B^0)^2 - (\mathbf{a} + \mathbf{b})^2 \\ &\geq (A^0 + B^0)^2 - (|\mathbf{a}| + |\mathbf{b}|)^2 \\ &\geq ((A^0)^2 - |\mathbf{a}|^2) + ((B^0)^2 - |\mathbf{b}|^2) + 2A^0B^0 - 2|\mathbf{a}||\mathbf{b}| \\ &> 0 \ , \end{aligned} \tag{154}$$

where I used the triangle inequality $|\mathbf{a} + \mathbf{b}| \ \leq |\mathbf{a}| + |\mathbf{b}|$ and that $A^0 > |\mathbf{a}|$ and $B^0 > |\mathbf{b}|$. If you did optional question 1 on the first problem sheet, you can also get this result directly from that.

Now onto the main proof.

Proof: The 4-momenta of the particles are all time-like 4-vectors with positive time-components. Therefore, the total 4-momentum has a positive time-component, and (by the lemma above) is time-like. By the result in Section 9, there hence exists a frame in which it has zero spatial component, which is the COM frame.

The centre of momentum frame will be very useful when we study particle collisions. Particularly important is the total energy in the COM frame, E_{COM}. This is typically most easily calculated using invariants.

Centre of momentum frame energy from invariants

From Eq. (153), $E_{\rm COM}^2 = c^2 P_{\rm tot} \cdot P_{\rm tot}$. Being invariant, $P_{\rm tot} \cdot P_{\rm tot}$ can be evaluated in any convenient frame, e.g. the lab frame.

E.g., for a system of two massive particles of masses m_1 and m_2 and 4-momenta P_1 and P_2,

$$\begin{aligned} E_{\rm COM}^2 &= c^2 \left(P_1 + P_2\right) \cdot \left(P_1 + P_2\right) \\ &= m_1^2 c^4 + m_2^2 c^4 + 2 P_1 \cdot P_2 c^2 \ , \end{aligned} \tag{155}$$

So $E_{\rm COM}$ can be found by evaluating $P_1 \cdot P_2$ in whatever frame we have information.

Example: Energy in centre of momentum frame

Consider two protons of mass $0.94\,{\rm GeV}/c^2$. Find the total energy in the COM frame if both have energy 5 TeV and equal and opposite velocities in the lab frame. What is the total energy in the COM frame if, instead, if one has energy of 10 TeV and the other is stationary in the lab frame?

In the first case the lab frame is the COM frame so $E_{\rm COM} = 10\,{\rm TeV}$.

In the second case, the 4-momenta of the two particles are $P_1 = (mc, 0, 0, 0)^T$ and $P_2 = (E_2/c, p_2, 0, 0)^T$. Therefore, using Eq. (155),

$$\begin{aligned} & E_{\rm COM}^2 = 2m^2c^2 + 2mE_2c^2 \\ \implies \ & E_{\rm COM} \simeq \sqrt{2mc^2 E_2} \\ & \simeq 0.14\,{\rm TeV} \ , \end{aligned} \tag{156}$$

where we used that $m \ll E_2$. Evidently the COM energy is much higher with symmetric beams. As will be clear in the next Section, this is why modern particle colliders use colliding beams rather than fixed targets, despite the engineering challenges that this leads to.

Example: Energy in different frames

Suppose that in the lab frame two electrons (each of mass $m_e = 0.51\,{\rm MeV}/c^2$) have opposite velocities such that the COM energy is 100 GeV. What is the energy of one of the electrons in the frame S' in which the other is at rest?

One approach is to use invariants. In S' we have $P_{\rm tot} = (E_1'/c + m_e c, p_1', 0, 0)^T$, where E_1' and p_1' are the energy and momentum of the moving electron. Hence

$$\begin{aligned} E_{\rm COM}^2 &= (E_1' + m_e c^2)^2 - p_1'^2 c^2 \\ &= (E_1'^2 - p_1'^2 c^2) + 2E_1' m_e c^2 + m_e^4 c^4 \\ &= 2E_1' m_e c^2 + 2m_e^2 c^4 \ . \end{aligned} \tag{157}$$

Since $E_{\rm COM} = 100\,{\rm GeV} \gg m_e c^2$, we have $E_1' \simeq E_{\rm COM}^2/(2m_e c^2) = 2.0 \times 10^7\,{\rm GeV}$.

The same result can alternatively be obtained by first finding the γ factor of the electrons in the lab frame from $E = \gamma mc^2$, and then Lorentz transforming the 4-momentum of one electron into the rest frame of the other.

Sometimes it can be useful to know the velocity $\mathbf{v}_{\rm COM}$ of the COM frame relative to the lab frame. This can be found using

$$\mathbf{v}_{\rm COM} = \frac{\mathbf{p}_{\rm tot} c^2}{E_{\rm tot}} \ , \tag{158}$$

where the quantities on the right-hand side are evaluated in the lab frame (this is the same expression as for an individual particle, Eq. (130)).

Proof: If the momenta of all the particles are aligned in one direction, e.g. along the x-axis, Eq. (158) can be derived from the Lorentz boost. In the COM frame, S' say, $P'_{\rm tot} = (E_{\rm COM}/c, 0, 0, 0)^T$, while in the lab frame $P_{\rm tot} = (E_{\rm tot}/c, p_{\rm tot}, 0, 0)^T$. Hence, the COM frame is moving relative to the lab frame in the x-direction with $v_{\rm COM}$ such that

$$\begin{pmatrix} E_{\rm COM}/c \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_{v_{\rm COM}} & \frac{-v_{\rm COM}}{c}\gamma_{v_{\rm COM}} \\ \frac{-v_{\rm COM}}{c}\gamma_{v_{\rm COM}} & \gamma_{v_{\rm COM}} \end{pmatrix} \begin{pmatrix} E_{\rm tot}/c \\ p_{\rm tot} \end{pmatrix} \ . \tag{159}$$

Solving for $v_{\rm COM}$ from the spatial components leads to Eq. (158).

If the momenta are not all aligned, first rotate the spatial coordinate axes so that $\mathbf{p}_{\rm tot}$ is in the x-direction. Then, the COM frame is obtained by boosting in the (post-rotation) x-direction as above. Rotating back to the original spatial coordinate axes leads to Eq. (158).

20 Relativistic collisions and decays

Relativistic energy and momentum conservation plays a crucial role in the decay and scattering of elementary particles. The study of such processes has been *the* key experimental route to learning about fundamental particle physics.

As discussed in Section 17, a crucial fact is that the total rest mass of a system is not conserved and instead only total 4-momentum is conserved (throughout we assume there are no external forces). As a result, processes occur in which the incoming and outgoing particles do not have the same total rest masses. This leads to a much greater diversity of particle interactions than would otherwise be possible. Some features of such processes can be inferred simply from conservation of energy and momentum without knowledge of the underlying interactions that lead to the e.g. scattering or decay.

Of course, ultimately we want to know the microphysics that determines the interactions between elementary particles, which allows the rate of decay and scattering processes to be predicted from first principles. However, with the exception of elastic electromagnetic interactions, this is

complicated and you'll have to wait until the 4th year course on QFT (or read ahead) to see how it's done. In fact, in some cases (e.g. scattering of protons) even though we are convinced that we know the underlying theory we cannot (yet?) make precise first principles predictions for many processes we care about. Instead, we still need to use inputs from experiment.

In the case of a particle decaying, some of its rest mass is converted into kinetic energy of the decay products (strictly speaking, the total rest mass after the decay could precisely equal that before, but in practice this never happens).

For collisions there are several possibilities.

Classifying collisions

We classify collision processes according to how the total mass before m_{before} and after m_{after} compare.

- Elastic collisions: $m_{\text{after}} = m_{\text{before}}$ so total kinetic energy is conserved.
- Explosive collision: $m_{\text{after}} < m_{\text{before}}$ so total kinetic energy increases.
- Sticky collision: $m_{\text{after}} > m_{\text{before}}$ so total kinetic energy decreases.

You can think of a decay as an explosive collision with an initial state of only a single particle.

Side note: I don't remember actually seeing the terminology explosive or sticky collisions used in the theoretical physics research literature (normally I'd just call both just "inelastic"). But these seem to be the words people use when teaching special relativity, so let's go with it.

20.1 Decays

Suppose a free particle of rest mass $m_A > 0$ decays into particles of mass m_i. We can study this in any frame we like, so let's use the rest frame of the initial particle. Then the initial 4-momentum is $(m_A c, 0, 0, 0)^T$ and the final 4-momentum is $\sum P_i$. Since $P_i^0 \geq m_i$, conservation of the time component of 4-momentum requires that

$$m_A \geq \sum m_i \ , \tag{160}$$

with the equality satisfied only if the speed of all the decay products is zero. Provided that Eq. (160) is satisfied there exist P_i such that 4-momentum is conserved (this can be shown by slightly generalising the examples below), so the decay is *kinematically* allowed.

Even if 4-momentum conservation allows a decay, it might not occur due to the microscopic physics. A simple example is that electric charge is conserved, forbidding many decay processes. A more subtle example is that we have never observed proton decay $p \to e^+\gamma\gamma$ even though it is

kinematically allowed. The reason for this is understood to be the conservation of baryon number, which is an accidental consequence of the structure of the Standard Model of particle physics.

A neat argument rules out the decay of an isolated massless particle (e.g. a photon) into massive particles. As seen in Section 19, for such a final state there is a COM frame in which $P_{\rm tot} = (E_{\rm COM}/c, 0, 0, 0)^T$. However, since no frame exists in which a massless particle has zero spatial momentum, such a decay cannot conserve 4-momentum. The same argument rules out decay to two massless particles that are not collinear (the case of a photon decay to two photons travelling in the same direction is actually forbidden by angular momentum conservation).

Example: Particle Decay I

Consider an unstable particle of type A, of rest mass m_A, that decays by splitting into two particles of type B (labelled 1,2), both of rest mass m_B. This process can be represented as

$$A \to BB \ , \tag{161}$$

(which is more like the notation for a chemical reaction than a genuine equation). Calculate the magnitude of the 3-velocities $|\mathbf{v}_{1,2}|$ of the produced particles in the rest frame of A.

Conservation of the spatial components of 4-momentum in the rest frame of A gives $\mathbf{0} = m_B\gamma_{v_1}\mathbf{v}_1 + m_B\gamma_{v_2}\mathbf{v}_2$, which implies $\mathbf{v}_1 = -\mathbf{v}_2$, so $|\mathbf{v}_1| = |\mathbf{v}_2| =: v$. Conservation of the time component of 4-momentum then gives $m_A = 2m_B\gamma_v$ so

$$v = c\sqrt{1 - \frac{4m_B^2}{m_A^2}} \ . \tag{162}$$

Not surprisingly, if $2m_B > m_A$ there is no solution for v and the decay cannot occur. In the limit $m_B/m_A \to 0$ we have $v/c \to 1$.

Consider now the lab frame in which A is initially moving at speed w and suppose that the two B are produced with velocities aligned and anti-aligned with the direction of motion of A. What are the velocities v' and energies of the two B particles in this frame?

The simplest way to find the velocities is the velocity addition formula. Combined with Eq. (162), this gives

$$v' = \frac{w \pm c\sqrt{1 - \frac{4m_B^2}{m_A^2}}}{1 \pm \frac{w}{c}\sqrt{1 - \frac{4m_B^2}{m_A^2}}} \tag{163}$$

where the $+$ and $-$ signs correspond to the particle produced travelling forward and backward in A's rest frame respectively. The energies of the particles are given by

$$E = \gamma_{v'} m_B c^2 \ , \tag{164}$$

with the values of v' from Eq. (163). After some algebra

$$E = \frac{(m_A + \frac{w}{c}\sqrt{m_A^2 - 4m_B^2})c^2}{2\sqrt{1-\frac{w^2}{c^2}}} \ . \tag{165}$$

An alternative, probably nicer, way to get the energies is to use that the 4-momenta of the produced particles in A's rest frame are $P_{1,2} = (m_A c/2, \pm\sqrt{m_A^2 c^2/4 - m_B^2}, 0, 0)^T$ (using that $P_{1,2}^2 = m_B^2 c^2$). Lorentz transforming to the lab frame gives the time-like component of the 4-momentum in this frame $P' = E'/c$, so

$$E' = \gamma_w \frac{m_A c^2}{2} \pm w\gamma_w \sqrt{m_A^2 c^2/4 - m_B^2} \ , \tag{166}$$

which is the same result as above.

Example: Particle Decay II

More generally, suppose a particle A decays to two particles of type B and C with different masses. Working in the rest frame of A, conservation of the spatial and time parts of 4-momentum gives

$$\begin{aligned} \mathbf{0} &= \mathbf{p}_B + \mathbf{p}_C \ , \\ m_A &= \sqrt{|\mathbf{p}_B|^2 + m_B^2} + \sqrt{|\mathbf{p}_C|^2 + m_C^2} \ , \end{aligned} \tag{167}$$

where $\mathbf{p}_B = m_B \gamma_B \mathbf{v}_B$ etc. Eliminating $\mathbf{p}_C$, some algebra *(exercise)* gives

$$\gamma_B = \frac{m_A^2 + m_B^2 - m_C^2}{2m_A m_B} \ , \tag{168}$$

and

$$\gamma_C = \frac{m_A^2 + m_C^2 - m_B^2}{2m_A m_C} \ . \tag{169}$$

Expressions for v_B and v_C can be obtained from γ_B and γ_C, although the results are not particularly insightful.

In realistic decays we often do not have information about the energy and momentum of one of the particles, either the initial particle (e.g. if this is itself produced from a collision) or one of the decay products (e.g. a neutrino that escapes the detector). In these situations it's often a good idea to isolate the unknown 4-momentum, P say, on one side of the 4-momentum conservation equation. Squaring both sides eliminates the unknown energy and momentum because $P \cdot P = m^2 c^2$. This trick is illustrated in the following examples.

Example: Reconstructing initial mass

Suppose that a new particle, of type X, is produced in a particle collider and quickly decays to known particles that are detected. In the process the energies and momenta of the known particles are all measured. The X particle is typically not produced at rest in the lab frame. Can we determine its mass m_X?

The answer is yes. Conservation of 4-momentum gives $P_X = \sum_i P_i$ where i labels the decay products. Squaring this equation

$$m_X^2 c^2 = \left(\sum_i P_i\right) \cdot \left(\sum_i P_i\right) \ . \tag{170}$$

Since the P_i are measured, we can easily find m_X. E.g. if there are two particles produced

$$\begin{aligned} m_X^2 c^2 &= P_1^2 + P_2^2 + 2P_1 \cdot P_2 \\ &= m_1^2 c^2 + m_2^2 c^2 + 2E_1 E_2/c^2 - 2\mathbf{p}_1.\mathbf{p}_2, \end{aligned} \tag{171}$$

in obvious notation. Note that we are essentially just calculating the centre of momentum energy of the produced particles.

Example: Higgs Decay

Analysis of data from the ATLAS and CMS detectors at the LHC discovered the Higgs boson, with mass $m_h = 125\,\mathrm{GeV}/c^2$. The first signal of this was through searches for its decays into two photons γ, i.e. $h \to \gamma\gamma$. What is the energy and 4-momenta of the photons produced in the rest frame of the Higgs?

In the rest frame of the Higgs the photons have energy $m_h c^2/2$ by conservation of energy. They can be emitted in any direction, but must be back to back, so have 4-momentum P_1 and P_2 of

$$\begin{aligned} P_1 &= \frac{m_h c}{2}\,(1, \hat{\mathbf{n}}) \\ P_2 &= \frac{m_h c}{2}\,(1, -\hat{\mathbf{n}}) \ , \end{aligned} \tag{172}$$

where $\hat{\mathbf{n}}$ is any unit 3-vector.

Consider now the lab frame, in which the Higgs energy $E_h \geq m_h c^2$. If one of the photons is measured to have energy E_γ then what is angle between the photon and the Higgs direction of travel? Working in the limit $E_h \gg m_h c^2$, what is the largest and smallest energy that the photons can have in the lab frame as a function of E_h?

Call the 4-momentum of the measured photon in the lab frame $P_\gamma = E_\gamma/c(1, \hat{\mathbf{m}})$ where $\hat{\mathbf{m}}$ is a unit vector, and the 4-momentum of the second photon in this frame P'_γ. Then 4-momentum

conservation gives $P_h = P_\gamma + P'_\gamma$. Using the trick above to eliminate P'_γ,

$$\begin{aligned} P'_\gamma &= P_h - P_\gamma \\ \implies 0 &= P_h \cdot P_h + P_\gamma \cdot P_\gamma - 2P_h \cdot P_\gamma \\ \implies 0 &= m_h^2 c^2 - 2P_h \cdot P_\gamma \ . \end{aligned} \tag{173}$$

Writing the 4-momentum of the Higgs $P_h = (E_h/c, \mathbf{p}_h)$, this implies

$$\begin{aligned} 2E_h E_\gamma - 2E_\gamma |\mathbf{p}_h| c\cos\theta &= m_h^2 c^4 \\ \implies \cos\theta &= \frac{2E_h E_\gamma - m_h^2 c^4}{2E_\gamma \sqrt{E_h^2 - m_h^2 c^4}} \ , \end{aligned} \tag{174}$$

where θ is the angle between the photon's and the Higgs' directions.

Solving for E_γ

$$E_\gamma = \frac{m_h^2 c^4}{2E_h \left(1 - \cos\theta \sqrt{1 - m_h^2 c^4 / E_h^2}\right)} \ . \tag{175}$$

Note that a solution for E_γ exists for any $\cos\theta$.

E_γ is maximised when $\cos\theta = 1$ (i.e. the photon is produced in the forward direction) with $E_\gamma \simeq E_h$ for $E_h \gg m_h c^2$, and it is minimised for $\cos\theta = -1$ (i.e. produced backwards) with $E_\gamma \simeq m_h^2 c^4/(4E_h)$ for $E_h \gg m_h c^2$. This is physically reasonable. The photon emitted forward carries the majority of the energy and momentum, while the backwards photon just compensates for the difference between the dispersion relations $E_h^2 = |\mathbf{p}_h|^2 c^2 + m_h^2 c^4$ and $E_\gamma^2 = |\mathbf{p}_\gamma|^2 c^2$.

The same result can be obtained by Doppler shifting the two photons from the frame in which the Higgs is at rest to the lab frame.

Example: Spontaneous emission

A very similar calculation arises in the decay of an excited atom to its ground state by emission of a photon. Suppose the initial atom is stationary with mass $M_\star$, and the ground state has mass M. What is the energy of the emitted photon?

Call the 4-momentum of the initial excited atom $P_\star = (M_\star c, 0, 0, 0)^T$, that of the final ground state atom $P = (E/c, \mathbf{p})^T$, and that of the photon P_γ. 4-momentum conservation gives $P_\star = P + P_\gamma$. Note that the ground state atom recoils to conserve momentum. We eliminate

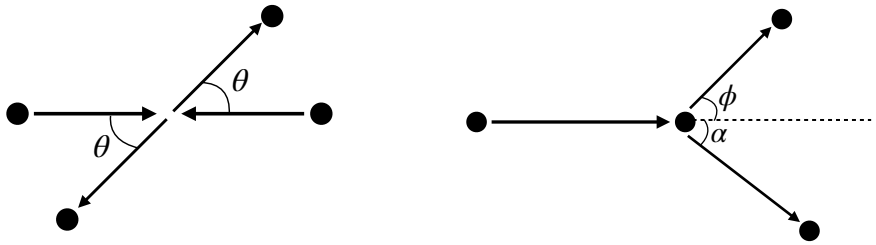


Figure 10: The kinematics of elastic scattering of two identical particles, in the centre of momentum frame (left) and the lab frame with one of the particles initially stationary (right).

the unknown 4-momentum of the ground state atom

$$\begin{aligned} P &= P_\star - P_\gamma \\ \implies P^2 &= P_\star^2 + P_\gamma^2 - 2P_\star \cdot P_\gamma \\ \implies M^2c^2 &= M_\star^2 c^2 - 2M_\star E_\gamma \\ \implies E_\gamma &= \frac{(M_\star^2 - M^2)\,c^2}{2M_\star} \,. \end{aligned} \tag{176}$$

20.2 Elastic collisions

As an example of an elastic collision we analyse the scattering of two identical particles, protons p say, without the production of any additional particles $pp \to pp$. We do so in two reference frames: the COM frame and a frame in which one of the protons is initially stationary, which we suppose is the lab frame.

In the COM frame the kinematics are simple. Call the initial 4-momenta in this frame $P_{1,2}$ and the final 4-momenta $P_{3,4}$, with $P_1 = (E_1, \mathbf{p}_1)$ etc. By definition $\mathbf{p}_1 + \mathbf{p}_2 = 0$. We rotate our coordinate system such that proton 1 is travelling along the x-axis, so

$$P_1 = (m\gamma_v c, m\gamma_v v, 0, 0)^T \,, \qquad P_2 = (m\gamma_v c, -m\gamma_v v, 0, 0)^T \,, \tag{177}$$

where $v = |\mathbf{v}_1| = |\mathbf{v}_2|$. Conservation of the spatial part of the total 4-momentum requires that $\mathbf{p}_3 = -\mathbf{p}_4$ and hence conservation of energy fixes that the speed of the two outgoing protons is also v. The only unfixed quantity is the direction at which the outgoing protons travel. Let's use the remaining freedom in our choice of axis such that they travel in the x-y plane, so

$$P_3 = (m\gamma_v c, m\gamma_v v\cos\theta, m\gamma_v v\sin\theta, 0)^T \,, \qquad P_4 = (m\gamma_v c, -m\gamma_v v\cos\theta, -m\gamma_v v\sin\theta, 0)^T \,. \tag{178}$$

The situation is illustrated in Fig. 10 left.

The scattering is more interesting in the lab frame (illustrated in Fig. 10, right). We denote quantities in this frame by $'$ and align our coordinate system so the incoming proton is travelling along the x-axis.

(For comparison, note that in Newtonian mechanics the two particles would separate with angle $90°$ *in the lab frame, or forward scatter with one particle finishing stationary. Conservation of energy and momentum means* $v_1'^2 = v_3'^2 + v_4'^2$ *and* $\mathbf{p_1}' = \mathbf{p_3}' + \mathbf{p_4}' \implies |\mathbf{p}_1'|^2 = |\mathbf{p}_3'|^2 + |\mathbf{p}_4'|^2 + 2\mathbf{p}_3'.\mathbf{p}_4'$*, which together imply* $\mathbf{p}_3'.\mathbf{p}_4' = 0$*.)*

Call the speed of the incoming proton in the lab frame v and assume it is travelling in the positive x-direction. Then, using Eq. (158), the speed of the COM frame relative to the lab frame is

$$v_{\text{COM}} = \frac{|\mathbf{p}_{\text{tot}}|c^2}{E_{\text{tot}}} = \frac{\gamma_v v}{1+\gamma_v} = \frac{v}{\sqrt{1-v^2/c^2}+1} \ . \tag{179}$$

As expected $v_{\text{COM}} < v$ and $v_{\text{COM}} \to v/2$ for $v/c \to 0$. The proton that is stationary in the lab frame has speed v_{COM} in the COM frame. Since in the COM frame the protons have equal and opposite velocities, the other proton has speed v_{COM} in the COM frame too (you can check this with the velocity addition formula Eq. (81)).

In the COM frame, after the collision the protons have 3-velocities of $\mathbf{v}_3' = (v_{\text{COM}}\cos\theta, v_{\text{COM}}\sin\theta, 0)$ and $\mathbf{v}_4' = (-v_{\text{COM}}\cos\theta, -v_{\text{COM}}\sin\theta, 0)$.

One way to get the velocities of the particles in the lab frame after scattering is by applying the general velocity addition formula, Eq. (87), to the proton's velocities in the COM frame after the collision. The relative speed of the two frames is v_{COM}. Therefore, the protons' velocity in the lab frame is

$$\begin{aligned} \mathbf{v}_3 &= \left(\frac{v_{\text{COM}}(1+\cos\theta)}{1+v_{\text{COM}}^2\cos\theta/c^2}, \frac{v_{\text{COM}}\sin\theta}{\gamma_{v_{\text{COM}}}(1+v_{\text{COM}}^2\cos\theta/c^2)}, 0 \right)^T \\ \mathbf{v}_4 &= \left(\frac{v_{\text{COM}}(1-\cos\theta)}{1-v_{\text{COM}}^2\cos\theta/c^2}, \frac{-v_{\text{COM}}\sin\theta}{\gamma_{v_{\text{COM}}}(1-v_{\text{COM}}^2\cos\theta/c^2)}, 0 \right)^T \ . \end{aligned} \tag{180}$$

From Eq. (180), the angles of the outgoing protons' trajectories to the x-axis, ϕ and α (as defined in Fig. 10, right), satisfy

$$\begin{aligned} \tan\phi &= \frac{\sin\theta}{\gamma_v(1+\cos\theta)} \\ \tan\alpha &= \frac{-\sin\theta}{\gamma_v(1-\cos\theta)} \ , \end{aligned} \tag{181}$$

where θ is the scattering angle in the COM frame (in our conventions $\alpha < 0$ in the sketch in Fig. 10 right). Therefore, the angle between the two protons $\phi - \alpha$ is given by

$$\tan(\phi-\alpha) = \frac{\tan\phi - \tan\alpha}{1+\tan\phi\tan\alpha} = \frac{2\gamma_v}{(\gamma_v^2-1)\sin\theta} \ . \tag{182}$$

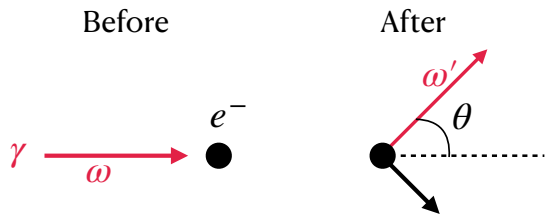


Figure 11: Compton scattering in the frame in which an electron is initially stationary. An incoming photon of angular frequency ω scatters off the electron, which recoils, leading to the photon angular frequency decreasing to ω'.

At fixed non-zero $\sin\theta$, for small velocities $\gamma_v \simeq 1$ and the angle between the particles is approximately 90° in the lab frame. For larger γ_v, the angle between the outgoing particles is noticeably smaller than 90° in the lab frame, in contrast to the Newtonian prediction.

An alternative way to derive Eq. (182) is by Lorentz transforming the protons' 4-velocity (or equivalently 4-momenta) from the COM frame to the lab frame.

20.3 Compton scattering

Compton scattering is the scattering $e^-\gamma \to \gamma e^-$ of a photon γ off an electron e^- resulting in the photon's frequency changing (analogous processes occur for a photon and any other charged particle). Given our definitions, Compton scattering is elastic because the total rest mass is unchanged.[16]

Usually we are interested in Compton scattering in the frame that the electron is initially stationary and recoils after scattering. This situation is shown in Fig. 11. We define the incoming photon to be travelling in the direction $\hat{\mathbf{n}}$ and have angular frequency ω such that its 4-momentum

$$P_\gamma = \frac{\hbar\omega}{c}(1, \hat{\mathbf{n}})^T \ . \tag{183}$$

Likewise, the outgoing photon travels in the direction $\hat{\mathbf{n}}'$ and has angular frequency ω' with 4-momentum

$$P'_\gamma = \frac{\hbar\omega}{c}(1, \hat{\mathbf{n}}')^T. \tag{184}$$

The initial and final 4-momenta of the electron are

$$P_e = (m_e c, 0, 0, 0)^T \ , \qquad P'_e = (m_e\gamma_{v'}c, m_e\gamma_{v'}\mathbf{v}')^T \ , \tag{185}$$

where $\mathbf{v}'$ is the outgoing electron velocity.

[16]Note however, different people define elastic scattering in slightly different ways.

Usually we want to determine the outgoing photon's frequency as a function of the incoming photon's frequency and its scattering angle θ (i.e. the angle between the incoming and outgoing photon). Therefore, we rearrange conservation of 4-momentum to $P'_e = P_\gamma + P_e - P'_\gamma$ and square to eliminate the unknown P'_e. This gives

$$\begin{aligned} m_e^2c^2 &= P_\gamma^2 + P_\gamma'^2 + P_e^2 + 2P_e \cdot (P_\gamma - P'_\gamma) - 2P_\gamma \cdot P'_\gamma \\ &= m_e^2c^2 + 2m_e\hbar(\omega - \omega') - \frac{2\hbar^2}{c^2}\omega\omega'(1 - \hat{\mathbf{n}}'.\hat{\mathbf{n}}) \ . \end{aligned} \tag{186}$$

Simplifying and using that $\hat{\mathbf{n}}'.\hat{\mathbf{n}} = \cos\theta$

$$m_ec^2(\omega - \omega') = \hbar\omega\omega'(1 - \cos\theta) \ , \tag{187}$$

as required. This expression is slightly nicer in terms of the photon wavelengths $\lambda = 2\pi c/\omega$:

$$\lambda' - \lambda = \frac{h}{m_ec}(1 - \cos\theta) \ . \tag{188}$$

We recognise the prefactor $h/(m_ec)$ on the right-hand side as the Compton wavelength of the electron (explaining its name).

Since the energy of a photon $E = \hbar\omega = hc/\lambda$, a photon scattering off a stationary electron loses energy if $\theta \neq 0$, with the maximum loss for backward scattering $\theta = \pi$, and no energy change for forward scattering $\theta = 0$. This is not surprising since the electron gains energy for $\theta \neq 0$.

The scattering of a photon off a moving electron can be analysed similarly. In fact this can be done by transforming the results above to a frame in which the electron is initially moving. In such a frame, the photon energy can either increase or decrease. This is sometimes called inverse Compton scattering, although it's the same process.

Note that, in the absence of interactions with other particles, $e^-\gamma \to e^-$ is impossible: in the rest frame of the final electron, energy conservation is violated (in contrast, spontaneous emission from an atom is possible because the rest mass of the atom changes). Likewise $e^- \to e^-\gamma$ is forbidden by an analogous argument. Eq. (187) confirms this: it cannot be satisfied if one of the incoming or outgoing photon energies is zero.

20.4 Inelastic collisions

Sticky collisions are particularly important for particle physics because they allow new particles with large masses to be discovered by colliding the relatively light particles that we can do experiments with. Let's start with a theorem that isn't very surprising but is still worth being careful about.

Theorem: The total mass of particles after a collision cannot exceed the initial energy in the centre of momentum frame. Moreover, if there are at least two particles after the collision and the sum of their mass energy is less than or equal to the initial COM energy:

$$E_{\rm COM} \geq \sum_f m_f c^2 \ , \tag{189}$$

where f labels the particles after the collision, then there exists a set of 4-momentum P_f such that 4-momentum is conserved and the process is kinematically allowed. (As mentioned, even if a process is kinematically allowed it might not occur due to the details of the underlying interactions.)

Proof: Work in the centre of momentum frame, so the total 4-momentum before the collision $P_{\rm tot} = (E_{\rm COM}/c, 0, 0, 0)$. After the collision the energy component of the total 4-momentum is $\sum_f \gamma_f m_f c$. Since $\gamma_f \geq 1$ if Eq. (189) is not satisfied energy cannot be conserved. Meanwhile, if $E_{\rm COM} = \sum_f m_f c^2$ then 4-momentum is conserved by all the particles being at rest after the collision in the centre of momentum frame. Finally, it is easy to check that if $E_{\rm COM} > \sum_f m_f c^2$ then energy and momentum conservation can always be satisfied. For example, this can be done by picking any two particles in the final state and giving them equal and opposite 3-momenta of the appropriate magnitude.

Example: A new particle from proton-proton collisions

Suppose we collide two protons, each of mass m_p. After the collision, we hope to be left with these protons, together with a new particle X of mass m_X. Suppose our proton collider involves two beams of protons of equal energy pointed towards each other. What is the minimum lab frame energy E_i and gamma factor γ_i that the initial protons must have for this process to be possible?

The lab frame is the COM frame, so Eq. (189) requires that

$$E_{\rm COM} = 2E_i = 2\gamma_{v_i} m_p c^2 \geq 2m_p c^2 + m_X c^2 \ . \tag{190}$$

Hence, the minimum energy is $E_i = m_p c^2 + \frac{1}{2} m_X c^2$, and the corresponding $\gamma_{v_i} = 1 + m_X/(2m_p)$.

Suppose now that our collider fires a beam of protons at a target of stationary protons. What is the minimum lab-frame proton energy E_1 and gamma factor γ_1 required for the same process to occur?

As in the examples of Section 19 (see Eq. (157)), we calculate $E_{\rm COM}$ using invariants and the relation $P_{\rm tot}^2 = E_{\rm COM}^2$. In the lab frame $P_{\rm tot} = (m_p c + E_1/c, \mathbf{p}_1)^T$. Hence,

$$E_{\rm COM}^2 = (m_p c^2 + E_1)^2 - |\mathbf{p}_1|^2 = 2m_p c^2 E_1 + 2m_p^2 c^4 \ . \tag{191}$$

Eq. (189) then implies

$$E_1 \geq m_p c^2 + \frac{m_X^2}{2m_p} c^2 + 2m_X c^2 \ , \tag{192}$$

and

$$\gamma_1 \geq 1 + \frac{m_X^2}{2m_p^2} + \frac{2m_X}{m_p} \ . \tag{193}$$

From the example above (and anticipated in Section 19) we see that for $m_X \gg m_p$ much larger initial energy is needed for a collider with a single beam and a target than a symmetric collider. For a single beam, the required kinetic energy scales quadratically with m_X not linearly. This is simply because not all of the energy in the lab frame is available to be converted to rest mass, since the end products cannot be at rest due to momentum conservation.

Threshold energy

The minimum initial particle energy required for some collision process to occur, e.g. $pp \to ppX$ in the example above, is called the ***threshold energy***. Often this refers to the energy of a particle that collide with a stationary particle, but it can have different meanings in different contexts.

In reality, at the LHC the interactions are not between protons but between their constituent quarks and gluons. The energy available to make a new particle depends on the fraction of the proton's total energy that the particular quark or gluon involved in a collision carries. This is why constraints on new particles from the LHC only extend to masses of at most a few TeV despite the COM energy being 14 TeV. An electron or muon collider of the same COM energy would be much better for searching for new heavy particles because electrons and muons are not composite (as far as we know), but building colliders for these particles presents significant technical challenges.

Single particle final states

You might have noticed that in the theorem near Eq. (189) I specified that there are at least two particles after the collision. What happens if there is only a single particle after the collision, e.g. electron-positron annihilation to a new heavy particle $e^+e^- \to X$?

In this case the final particle must be stationary in the COM frame to conserve 3-momentum. Then, by energy conservation, the process can only occur if the COM energy is exactly correct, $E_{\text{COM}} = M_X$. This is like decay or spontaneous emission run backwards in time. However, rather than the final state momenta being produced with the right momenta, the initial state momenta must be "chosen" correctly. Consequently, such processes are relatively rare, although they do occur in some situations, e.g. an atom can absorb a photon leading to an excited state.

The reason such processes can occur is that because a particle with finite lifetime only approximately satisfies $E^2 - p^2c^2 = m^2c^4$, the initial particles' momenta do not need to be tuned with infinite precision. In more detail, by the uncertainty principle of quantum mechanics $\Delta E \Delta t \geq \hbar/2$, which means that if a particle has a lifetime τ its rest energy has uncertainty $\Delta E = \hbar/(2\tau)$.

Due to conservation of momentum, there is a mismatch between the frequency of the photon emitted when a stationary excited atom decays and the frequency of the photon that a stationary atom in the ground state would absorb. One of the questions on problem sheet 2 explores this and a related, experimentally important, phenomenon called the Mössbauer Effect.

Explosive collisions

The kinematics of explosive collisions can be analysed similarly to sticky collisions. An example of such a process is electron positron annihilation to two photons $e^+e^- \to \gamma\gamma$.[17] One crucial different to sticky collisions is that an explosive collision is kinematically allowed regardless of the kinetic energy of the initial particles (there is enough energy to produce the final state even if the initial particles are at rest). Instead there is a minimum total energy that the particles are produced with. For example, in $e^+e^- \to \gamma\gamma$ the sum of the energies of the two photons is at least $2m_ec^2$ in all frames.

20.5 Composite systems

So far we haven't talked much about the fact that the particles involved in decays and collisions are often not elementary, i.e. they have underlying structure. In reality, we know that e.g. nuclei are made of protons and neutrons, and protons and neutrons themselves are a complicated bound state of quarks and gluons. The mass of e.g. a nucleus, m, is not simply the sum of the rest masses of its constituents, but gets a contribution from the kinetic energies of its constituents (which are non-zero even in the rest frame of the nucleus) as well as from binding energies. None of our derivations relied on the particles being elementary, so the overall composite system, with 4-momentum P, still satisfies $P^2 = m^2c^2 = E^2/c^2 - |\mathbf{p}|^2$ and all our calculations still hold.

21 Newton's laws

Let's finally return to give the relativistic analogue of Newton's second law. As discussed at the start of Section 17, I will just state this.

[17] $e^+e^- \to \gamma$ is impossible because there exists a centre of momentum frame before the collision but not after.

Relativistic Newton's Law

$$\frac{dP}{d\tau} = F \ , \tag{194}$$

where P is the 4-momentum of a particle, τ is its proper time, and F is a 4-vector called the ***4-force***. Since both sides of Eq. (194) are 4-vectors, this equation is Lorentz covariant.

What are the components of F?

Given what we know from Newtonian mechanics, it's reasonable to postulate that the rate of change of 3-momentum with time $d\mathbf{p}/dt$ (where $\mathbf{p} = m\gamma v$ is the relativistic momentum) is related to a 3-vector representing the force $\mathbf{f}$ by

$$\frac{d\mathbf{p}}{dt} = \mathbf{f} \ . \tag{195}$$

In the way we've written it, this could almost be interpreted as a definition of $\mathbf{f}$. It has physics content once we have a Lagrangian, which shows that if the force comes from a potential then $\mathbf{f}$ is given by the gradient of this in the usual way, $\mathbf{f} = -\nabla V$.

Since $d/d\tau = \gamma d/dt$, the spatial part of $dP/d\tau$ is $\gamma d(\gamma m\mathbf{v})/dt = \gamma d\mathbf{p}/dt$. Combining Eqs. (194) and (195), this fixes that the spatial part of F is $\gamma\mathbf{f}$, i.e.

$$F = \begin{pmatrix} F^0 \\ \gamma\mathbf{f} \end{pmatrix} \ . \tag{196}$$

From Eq. (195), the time-like component of F is related to the rate of change of the particle's energy by

$$F^0 = \frac{1}{c}\frac{dE}{d\tau} = \frac{\gamma}{c}\frac{dE}{dt} \ . \tag{197}$$

The relation between the work done by a force and the change in a particle's energy, familiar from Newtonian physics, is also present in relativity. Differentiating both sides of $E^2 = m^2c^4 + |\mathbf{p}|^2c^2$ with respect to time gives

$$E\frac{dE}{dt} = 2\mathbf{p}.\frac{d\mathbf{p}}{dt}c^2 = 2\mathbf{p}.\mathbf{f}c^2 \ . \tag{198}$$

Using $E = \gamma mc^2$ and $\mathbf{p} = \gamma m\mathbf{u}$, gives

$$\frac{dE}{dt} = \mathbf{u}.\mathbf{f} \ , \tag{199}$$

just like in Newtonian mechanics, but with E the relativistic energy γmc^2.

Caution: most of the forces familiar to you from Newtonian mechanics do not remain valid in the relativistic regime and cannot simply be plugged into Eqs. (194) and (196) (e.g. Newtonian gravity must be replaced by general relativity). Moreover, many of the most interesting processes

in relativistic dynamics also involve quantum mechanics, and in this context Eq. (194) often isn't the most useful approach.

One situation in which Eq. (194) is useful is in analysing a charged particle moving in an electromagnetic field. This is not on the syllabus, but since it's a bit unsatisfying to not give any applications of Eq. (194) I give a brief overview below for the interested reader.

The force law for a particle of charge q moving in electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$ respectively, is

$$\frac{dP}{d\tau} = \frac{q}{c}\mathcal{F}U \ , \tag{200}$$

where U is the particle's 4-velocity and $\mathcal{F}$ is a matrix (in fact, a tensor)

$$\mathcal{F} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ E_x & 0 & cB_z & -cB_y \\ E_y & -cB_z & 0 & cB_x \\ E_z & cB_y & -cB_x & 0 \end{pmatrix} \ , \tag{201}$$

(in most books this tensor is called F, but here I use $\mathcal{F}$ to avoid confusion with the 4-force). The 4-force $F = \frac{q}{c}\mathcal{F}U$.

As an exercise you can check that the time-like component of Eq. (200) leads to the formula for the work done on a charged particle

$$\frac{dE}{dt} = q\mathbf{E}.\mathbf{u} \ , \tag{202}$$

where E is the particle's energy, $\mathbf{u}$ is its 3-velocity, and the spatial components lead to the Lorentz force law Eq. (17): $\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$.

Example: A particle in an electric field
A particle of charge q starts from rest in the lab frame at $x = 0$ at $t = 0$, and is accelerated by a uniform electric field E_x pointing in the x-direction. Find the particle's speed in the lab frame as a function of time.

The time and spatial components of Eq. (202) both simplify to

$$\frac{1}{(1 - v^2/c^2)^{3/2}}\frac{dv}{dt} = \frac{qE_x}{m} \ . \tag{203}$$

Integrating, and fixing the constant of integration, we obtain

$$\begin{aligned} \frac{v}{\sqrt{1 - v^2/c^2}} &= \frac{qE_x}{m}t \\ \implies v &= \frac{qE_x t}{m}\frac{1}{\sqrt{1 + q^2E_x^2t^2/(m^2c^2)}} \ , \end{aligned} \tag{204}$$

which reduces to the Newtonian expression for t such that $v \ll c$, and has $v \to c$ for $qE_x t/m \gtrsim c$. If required, the particle's position as a function of time can easily be found by integrating again.

Example: A particle in a magnetic field
A particle of charge q moves in a magnetic field $\mathbf{B} = (0, 0, B_z)$ with initial condition $\mathbf{v}_0 = (v_0, 0, 0)$ and position $(0, 0, 0)$ at $t = 0$ in the lab frame. Describe its trajectory.

The time and space components of Eq. (200) reduce to

$$\frac{dE}{dt} = 0 \ , \qquad \frac{d\mathbf{p}}{dt} = q\mathbf{v} \times \mathbf{B} \ . \tag{205}$$

The first of these implies E and therefore $|\mathbf{v}|$ and γ are constant. The second hence simplifies to

$$\frac{d\mathbf{v}}{dt} = \frac{q}{m\gamma}\mathbf{v} \times \mathbf{B} \ . \tag{206}$$

This solution to Eq. (206), with the given initial conditions, is circular motion in the x-y plane. This can be seen by writing out the components, differentiating to eliminate e.g. v_y, and integrating:

$$\begin{aligned} v_x &= v_0 \cos(\omega_c t) \\ v_y &= v_0 \sin(\omega_c t) \ , \end{aligned} \tag{207}$$

and

$$\begin{aligned} x &= \frac{v_0}{\omega_c} \sin(\omega_c t) \\ y &= \frac{v_0}{\omega_c} \left(1 - \cos(\omega_c t)\right) \ , \end{aligned} \tag{208}$$

with the angular frequency

$$\omega_c = \frac{qB}{\gamma m} \ . \tag{209}$$

The radius of the circle traced out

$$R = \frac{v_0}{\omega_c} = \frac{m\gamma v_0}{qB} = \frac{|\mathbf{p}|}{qB} \ . \tag{210}$$

For $v_0 \ll c$ the frequency and radius of the particle's circular motion reduce to the Newtonian expressions, but they differ substantially for $v_0 \simeq c$. This is important for particle colliders that make use of magnetic fields to keep particles moving in circular orbits while they are accelerated. The dependence of ω_c on γ in Eq. (209) means that if the particles are being accelerated by AC electric fields, the accelerating electric fields will go out of phase with the particles' motion as they get faster. In the simplest machines, known as cyclotrons, this limits the achievable speeds to $\simeq 0.1c$ for protons. This problem is addressed in synchrotrons (such as the LHC), which have magnetic fields that increase as the particles are accelerated so that ω_c and R stay constant.

22 Conclusions and what next?

With that, we have reached the end of our introduction to special relativity. Let me finish by commenting on a few notable omissions. Some of these could have been included with extra time, while others are more advanced and can only be taught later in your degree.

The most obvious shortcoming is that we have not shown that electrodynamics is Lorentz covariant, although the covariance of the wave equation Eq. (18) is a hint that we are on the right track. There is a beautiful formulation, hinted at in the last Section, that you will see in your third year. This reduces Maxwell's equations (taught to you next term) to

$$\partial_\mu \mathcal{F}^{\mu\nu} = J^\nu \ , \qquad \partial_\mu \tilde{\mathcal{F}}^{\mu\nu} = 0 \ , \tag{211}$$

where $\mathcal{F}$ is the 4×4 antisymmetric matrix containing the $\mathbf{E}$ and $\mathbf{B}$ fields given in Eq. (201) (usually called F), and $\tilde{\mathcal{F}}$ is a related matrix (usually called $\tilde{F}$). The asymmetry between the equations in Eq. (211) would vanish if magnetic monopoles exist (many theorists expect they do, but if so they are likely extremely heavy). Once you know index notation you will be able to see that these equations are Lorentz covariant.

Another major omission is the combination of special relativity and the Lagrangian formulation of mechanics (a few hints about this have been given in red comments). The Lagrangian formulation of special relativity for massive particles is certainly important, but arguably even more crucial is the formulation of electromagnetism. Continuing to show you equations without explaining them properly, Eq. (211) can be derived from a Lagrangian density

$$\mathcal{L} = -\frac{1}{4}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu} + J_\mu A^\mu \ , \tag{212}$$

and in fact even this Lagrangian can be derived from a principle known as gauge symmetry. Once special relativity is combined with quantum mechanics in QFT, gauge symmetry can be understood as a consequence of requiring a theory of massless spin-1 particles to respect causality and some other very reasonable conditions (probabilities sum to 1 etc.).

We could also have spent more time on the mathematical formalism underlying special relativity. One part of this is properly analysing the Lorentz group and its representations. This leads to the spectacular possibility of spin 1/2 particles. A closely related topic that we have also missed is proper definitions of 4-vectors, dual 4-vectors (elements of a space dual to that of 4-vectors similarly to what you will have seen in linear algebra), and more generally tensors.

Finally, another obvious next step is general relativity. This will be taught to you in the third year, but if you are enthusiastic there are many good books that should be accessible after working through these notes. For example the book "Relativity Made Relatively Easy" by Andrew Steane has a few chapters that get to interesting results, such as black holes, starting from the correct level.

A Time dilation and length contraction again

Here we show how time dilation and length contraction can be justified by considering particular physical setups, using the fact that light travels at speed c in all frames.

For time dilation, consider a light signal that, in a frame S, is sent in the y-direction and travels a distance L then returns. The time for this to occur is $\tau = 2L/c$. Meanwhile, in a frame S' moving relative to S with speed v in the x-direction, the light's journey takes a time t' during which the light travels a distance $\sqrt{(2L)^2 + (t'v)^2}$. Since the light has speed c

$$\begin{aligned} t' &= \frac{1}{c}\sqrt{(2L)^2 + t'^2 v^2} \\ \implies t' &= \frac{1}{\sqrt{1 - v^2/c^2}} \frac{2L}{c} \\ \implies t' &= \gamma\tau \ . \end{aligned} \tag{213}$$

If an observer in S uses the light's return as a clock, Eq. (213) is consistent with our general derivation of time dilation.

For length contraction, consider an observer in a frame S who sends a signal in the x-direction to a mirror a proper distance ℓ_0 away that is reflected back. Suppose the observer and mirror are stationary in S, so the light's out and back trip takes a time $\tau = 2\ell_0/c$. In a frame S' moving at velocity v relative to S in the x-direction, the light's trip takes a time

$$\begin{aligned} t' &= \frac{\ell'}{c+v} + \frac{\ell'}{c-v} \\ &= \frac{1}{1 - v^2/c^2} \frac{2\ell'}{c} \ , \end{aligned} \tag{214}$$

where ℓ' is the observer–mirror distance in S'. Since $t' = \gamma_v \tau$, we obtain $\ell' = \ell_0/\gamma_v$, which is length contraction.

B Examples of "paradoxes"

Some of the discussion in this Appendix follows the nice presentation by Prof. Taylor for a previous version of this course, available in original form at `https://users.physics.ox.ac.uk/~rtaylor/teaching.shtml`, and other parts follow Prof. David Tong's lecture notes. The calculations involved are a bit fiddly, so you may well have to work through them on pen and paper to follow all the details.

The twin paradox

Consider two twins, Alice and Bob say. Suppose Alice stays on Earth while Bob travels at speed v to a distant planet, and on arrival immediately turns around and comes back at the same speed. Define T to be the time that Bob arrives at the distant planet in the rest frame of the Earth, so his entire journey takes $2T$ in the Earth's rest frame. This is the amount Alice has aged when Bob gets back. The proper time along Bob's worldline is $2T/\gamma_v$, so he arrives back to find he is younger than Alice.

As yet this is simply time dilation. A "paradox" arises when you think about things from Bob's perspective. As far as he's concerned for almost the entire time he is happily sitting in an inertial reference frame, first watching Alice fly away and then watching her come back towards him. So from his perspective shouldn't Alice be younger? Why isn't the situation symmetric?

The resolution is that Alice remained in a single inertial frame throughout whereas Bob changes inertial frame in returning from the distant planet. Bob can, by doing experiments on his spaceship, determine that he has accelerated. So there is no symmetry between the two journeys and no reason to think the two will age the same amount. Aside from the fact that the change of reference frame is assumed to happen fast, the situation is no different to the two worldlines in Fig. 9, which start and end in the same place but with a longer proper time elapsed for the straight one.

To make sure that we're not tricking ourselves let's do a calculation (this is always a good idea!). Here I follow the presentation in Prof. David Tong's notes.

In Fig. 12 we draw the spacetime diagrams for Bob's journey with Alice's frame labelled S and Bob's two frames S' and S'' for the out and back parts respectively (left and right panels). Not all the angles in the diagram are precise but the axis labelling is such that $\gamma \simeq 3/2$, i.e. $v \simeq 3c/4$, with the planet a distance 2 away in Alice's frame (in whatever units you like).

On the way out, Bob's worldline is $x = vt$, equivalently $x' = 0$, while Alice sits at $x = 0$. Bob reaches the planet at the spacetime event E_1. Define the time coordinate of E_1 in S to be T, then E_1 has coordinates in S of $(ct, x) = (cT, vT)$, which is simultaneous with the event labelled Y. Doing a Lorentz transformation (or using the usual formula for time dilation) the coordinates of E_1 in the frame S' are $(ct', x') = (cT/\gamma, 0)$. In S' the event E_1 is simultaneous with X, which in S has coordinates $(ct, x) = (cT/\gamma^2, 0)$ (since Lorentz transforming this to S' we get $(ct', x') = (cT/\gamma, -vT/\gamma)$).

So far neither Alice or Bob has changed reference frame (we ignore the initial acceleration). As far as Alice is concerned Bob arrives at the planet at $ct = cT$ having only aged an amount T/γ, which is just normal time dilation. Likewise, in S' Bob arrives at the planet after T/γ, and at this time thinks that Alice has aged T/γ^2. So both see the other as having experienced time dilation by a factor γ as expected.

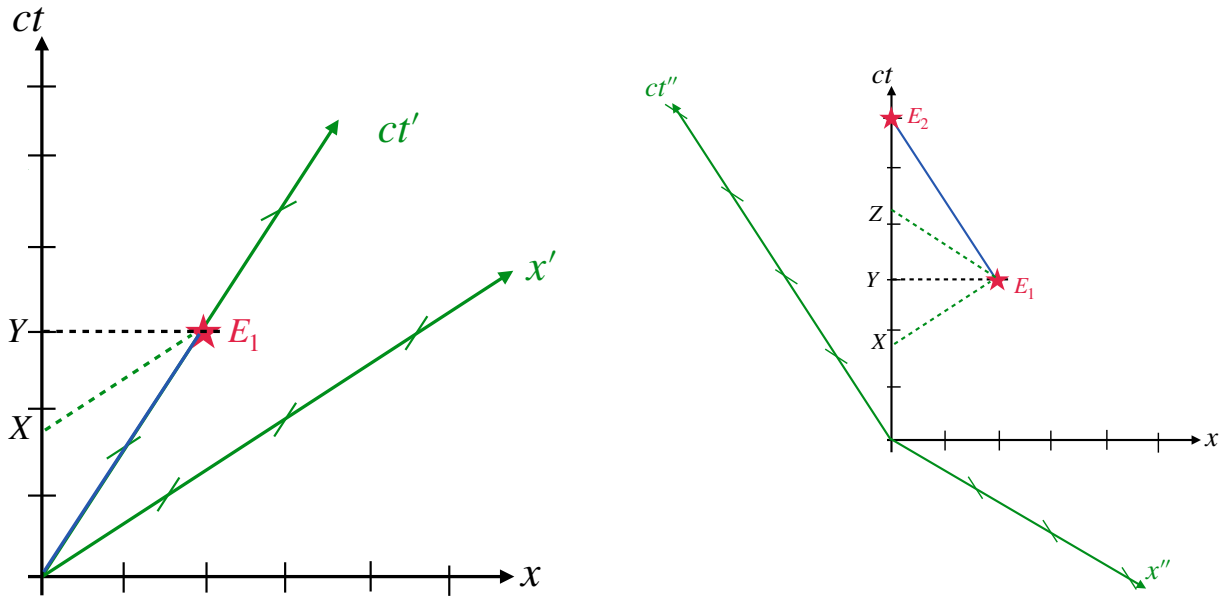


Figure 12: The twin "paradox" **Left:** The outward leg of Bob's journey (blue worldline) with S Alice's reference frame and S' Bob's during this leg. in S' the events X and E_1 are simultaneous, and in S the events Y and E_1 are simultaneous. **Right:** The return leg of Bob's journey, during which he's in the frame S''. In this frame Z and E_1 are simultaneous.

Now let's consider the return journey, but for the moment just think about a different person, Chris say, passing by Bob in the opposite direction just as Bob arrives at the planet. For Bob, the event E_1 is simultaneous with the event Z shown in the right panel of Fig. 12. Call the event when Chris arrives at Earth E_2. Then, going through a similar calculation to before, Chris ages by T/γ between E_1 and E_2 and sees that Alice has aged by T/γ^2. Meanwhile in Alice's frame E_1 was simultaneous with Y, so sees Chris ageing by T/γ as time dilation.

In this set-up both Bob on his outward journey and Chris on the return journey see Alice as again T/γ^2, so where did the missing time go such that Alice ages $2T$ when Bob makes a round trip? The answer is that in the time it takes for Bob to turn around and head home, Alice rapidly ages from X to Z as Bob's reference frame changes, and the event E_1 goes from being simultaneous with X to being simultaneous with Y. It would be a useful exercise to check that indeed the time separation between X and Z in Alice's frame is $2T - 2T/\gamma^2$.

To summarise, from Bob's perspective, Alice ages from 0 to X during his outward journey, from X to Z as he turns around, and from Z to $2T$ on his return journey.

Ladders and barns

A ladder of proper length $2L = 20\,\mathrm{m}$ approaches a barn of proper length $L = 10\,\mathrm{m}$ at a speed u such that $\gamma_u = 3$. An observer in the rest frame of the barn measures the ladder's length to be contracted by a factor of 3 and expects the whole ladder to fit in the barn. Meanwhile, an observer in the rest frame of the ladder measures the barn to be contracted by a factor of 3 and thinks that the ladder will not fit in the barn. Who is correct?

As usual, to reconcile observations in two different inertial reference frames we need to think more carefully about the question being asked. The most reasonable definition of a "ladder fitting into a barn" means the back end is inside the door before the front end hits the far wall. But since the time-ordering of space-like separated events varies between reference frames, the "before" in this definition suggests that observers might disagree (which isn't a problem given that there's no causality violation).

The spacetime diagram is drawn in Fig. 13, although not with correct angles or lengths for the numerical values above, so don't interpret the tick marks literally. The barn is at rest in the frame S, with entrance at x_1 and far wall at x_2, the ladder traces out the worldsheet in red, and is at rest in the frame S'. At $t = 0$ the front end of the ladder is at the entrance to the barn.

The front of the ladder hits the far end of the barn at event E_2, which has coordinates in S of $(ct_2, x_2) = (cL/u, x_2)$ corresponding to a time in S' of $t'_2 = \gamma_u L/u - \gamma_u u x_2/c^2$.

Meanwhile, in the barn's rest frame, at $t = 0$ the far end of the ladder is at $x_1 - 2L/\gamma$ due to length contraction.[18] Hence, it enters the barn at event E_1, which has coordinates in S of $(ct_1, x_1) = (c2L/(u\gamma), x_1)$, corresponding to a time in S' of $t'_1 = 2L/u - \gamma_u u x_1/c^2$.

Since $t_2 - t_1 = L/u - 2L/(u\gamma) > 0$ (with $\gamma_u = 3$ as above) in the barn's rest frame the back of the ladder enters the barn before the front hits the wall. Conversely,

$$\begin{aligned} t'_2 - t'_1 &= \gamma_u L/u - \gamma_u u x_2/c^2 - 2L/u + \gamma_u u x_1/c^2 \\ &= \gamma_u L/u - \gamma_u u L/c^2 - 2L/u \\ &< 0\ , \end{aligned} \tag{215}$$

where we used $x_2 - x_1 = L$, and on the final line plugged in numbers. So, as expected, in the ladder's rest frame the front of the ladder hits the far wall before the back of the ladder has entered.

Now let's ask what happens after the front of the ladder hits the wall at the far end of the barn. We assume that both the barn wall and the ladder are made of the most rigid material possible, the collision doesn't change the ladder's properties, and the front end stops at the end of

[18] If you want to avoid using the formula for length contraction, define two events with $t = 0$ one at the front and one at the back of the ladder, use the fact that the distance between them in S' is $2L$ and you will get the same result (this is basically just re-deriving length contraction).

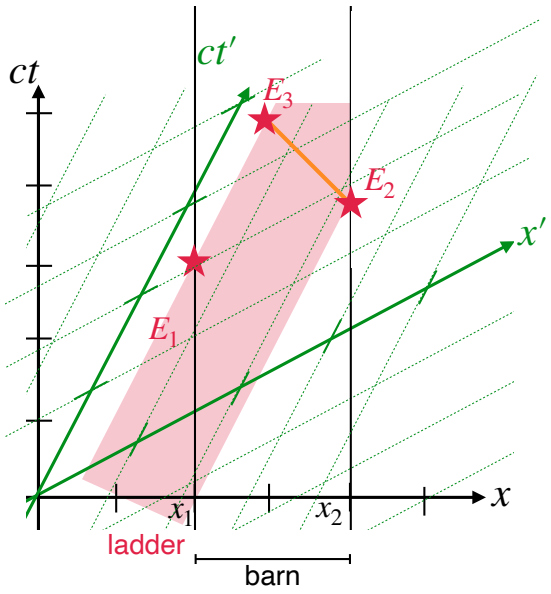


Figure 13: A ladder with worldsheet indicated by the red band (which is stationary in the frame S') going into a barn (stationary in the frame S) with entrance at x_1 and back wall at x_2. The front of the ladder hits the back wall at event E_2 and the back of the ladder enters the barn at event E_1. In the barn's rest frame E_2 is after E_1; the converse is true in the ladder's rest frame. The event E_3 is the first moment the back of the ladder can "know" that the front of the ladder has hit the wall.

the barn. The situation is potentially confusing since in the barn's reference frame the ladder is already entirely in the barn at this moment, whereas in the ladder's frame the back is still outside. Will the back have actually entered the tunnel or not? Where is its final resting place?

The simplest thing to clear up is the final situation: since both the ladder and barn finish at rest the final position of the back of the ladder must be outside the barn. But did it go in and bounce out? Or did it approach the end from outside and then stop? The answer must be the same in all frames, since it is simply the question "is there a spacetime event at which the back of the ladder is inside the barn?" which is a Lorentz invariant question.

The key thing is that information cannot travel faster than the speed of light. Therefore, regardless of how rigid the ladder is, the back of the ladder cannot know that the front has come to rest until the event E_3 on Fig. 13, which is the event at which a light signal emitted at E_2 reaches the back of the ladder. In the rest frame of the ladder, it takes a time $2L/c$ for the signal to reach the back of the ladder. Comparing the time for the back of the ladder to first "know" that the front has stopped with the time when it would enter the barn

$$\begin{aligned} t'_3 - t'_1 &= (t'_2 - t'_1) + (t'_3 - t'_2) \\ &= \gamma_u L/u - \gamma_u u L/c^2 - 2L/u + 2L/c \\ &> 0 \ . \end{aligned} \tag{216}$$

Therefore, even in the ladder's rest frame the back of the ladder enters the barn and then bounces out, which is consistent with what is found in the barn's rest frame.

Others

Other "paradoxes" that it might be interesting or useful to look at include:

- The rod-slit paradox, see A.P. French's book on special relativity, question 4.17
- Bell's spaceship paradox, see e.g. `https://math.ucr.edu/home/baez/physics/Relativity/SR/BellSpaceships/spaceship_puzzle.html`.