Chapter 4

Special Relativity

4.1 The speed of light and Einstein’s postulates

4.1.1 Before Einstein

Before Einstein, the nature of space and time was not generally thought to need discussion. Space and time were assumed to be completely different things. They were considered to be absolute, i.e., the same for all observers. Newton’s idea of time is summarised in this quotation from his “Principia”:

“Absolute, true and mathematical time, of itself and by its own true nature, flows uniformly on, without regard to anything external”.

(actually the fact that Newton wrote this down explicitly rather than taking it for granted is an example of his genius).

Light is a wave motion. There is no possible doubt about this, because light shows interference effects in experiments like Young’s slits. Other kinds of waves that we are familiar with propagate in a physical “medium”. There is some kind of material, whose oscillations constitute the wave motion.

Example: Sound waves travel through air, which is the physical material that oscillates. Air is compressible, and sound consists of variations of the density of air.

Example: Waves travel on the surface of water. The physical property that oscillates is the water level.

By analogy, it was thought in the 19th century that light must consist of the oscillations of some physical material, called the “luminiferous aether”. It was not known what this material was, but it was assumed that it must exist.

It is expected theoretically, and observed experimentally, that the observed speed of sound in air depends on the motion of the observer relative to the air.

For example, suppose a loudspeaker emits sound which is detected by an observer who is stationary with respect to the loudspeaker; the distance between them is \(d\). If the air is stationary with respect to the loudspeaker and observer, then the time taken for a pulse of sound to travel from the loudspeaker is \(d/v_0\), where \(v_0\) is the speed of sound in stationary
air. Now suppose the same observation is made in a strong wind, with the air moving at speed \( u \) in the direction from the loudspeaker to the observer, the distance between loudspeaker and observer being fixed, as before, at the value \( d \). Now a pulse of sound will travel more quickly, and will arrive in the shorter time \( d/(v_0 + u) \). If the wind blows in the opposite direction at speed \( u \), then the pulse will take the longer time \( d/(v_0 - u) \).

The wavelength of the sound is also changed by the motion of the air. If the loudspeaker emits sound at frequency \( f \), and the observer is stationary with respect to the loudspeaker, then the observer also hears the frequency \( f \). But \( \text{wavelength} = \text{speed}/\text{frequency} \). So with the wind blowing from the loudspeaker to the observer, the wavelength is

\[
\lambda = (v_0 + u)/f,
\]

and with the wind blowing from observer to loudspeaker it is

\[
\lambda = (v_0 - u)/f.
\]

The number of wavelengths that fit into the distance between loudspeaker and observer is

\[
d/\lambda = fd/(v_0 + u)
\]

in the first case, and

\[
fd/(v_0 - u)
\]

in the second case.

Therefore, if the aether exists, the speed of light should depend on the relative motion of the observer and the aether. And the wavelength should change correspondingly.

A major theoretical discovery in the 19th century was Maxwell’s equations, which gave a mathematical description of time-varying electric and magnetic fields. Maxwell’s equations predict the existence of “electromagnetic waves”, and also predict the speed of these waves. The predicted speed is the same as the observed speed of light. This implied that light consists of electromagnetic waves. But the derived speed comes straight out of Maxwell’s equations, which are differential vector equations, with no reference to either the coordinate system in which the vectors are defined, or to the motion of an observer relative to these coordinates!

### 4.1.2 The Michelson-Morley experiment

Nature presents us with a wonderful way of doing an experiment with light just like the experiment with the loudspeaker and the observer in a strong wind.

The Earth moves in a nearly circular orbit around the sun, at a speed of about \( 3 \times 10^4 \text{ m s}^{-1} \). So if there is a aether in the universe, presumably there is an aether “wind” rushing past us at this speed.

If a light source and an observer are fixed relative to the Earth, then the time taken for a light pulse to travel from the light source to the observer will be affected by the aether wind in the same way as in the loudspeaker example. Similarly, the number of wavelengths
that fit into the distance \( d \) between light source and observer will be \( fd/(c + u) \) when the aether wind is blowing from source to observer, but \( fd/(c - u) \) if it is blowing the other way, with \( c \) the speed of light relative to the aether and \( u \) the speed of the aether wind.

The Michelson-Morley experiment is designed to detect the aether wind using the ideas explained above. It is too difficult to measure the travel time of light pulses with sufficient accuracy, so instead the experiment uses interference effects to detect the changes of wavelength that should be caused by the aether wind.

![Schematic of the Michelson-Morley experiment.](image)

The picture shows a schematic of the experiment. Monochromatic light from a source strikes a “beam splitter” consisting of a half-silvered mirror at an angle of 45°. This splits the light into two beams which are reflected by two normal mirrors so that the beams recombine at the beam-splitter. The recombined beams are observed in a telescope. Because the light is monochromatic the two beams can interfere with each other.

To make it easier to observe interference, one of the normal mirrors is slightly tilted so as to produce interference fringes in the field of view. The aether wind effects are expected to differ for the two beams produced by the beam splitter. The apparatus is designed to be rotated. Rotation should change the aether-wind effects in the two arms, and this should be observed as a shift of the fringe pattern.

Numerical calculations (see the standard text books, e.g. Kleppner and Kolenkow, “An Introduction to Mechanics”, McGraw-Hill) show that the fringe shift should be easily observable. In spite of this, no effect at all is observed, within experimental error. The experiment has been repeated many times in the past 100 years, with greater and greater precision, but no effect has ever been detected.

**Conclusion:** the effects that would be expected from the aether theory are not observed. This casts serious doubt on the existence of the aether.
### 4.1.3 Einstein’s postulates

Einstein postulated that the aether does not exist. One reason is to ensure that Maxwell’s equations are equally valid for all observers, no matter what their relative velocity. This implies that all such observers observe the same velocity of light. According to Einstein, this is why the Michelson-Morley experiment produces a null result.

Einstein’s postulates are:

- “The laws of physics have the same form in all inertial systems.”

- “The velocity of light in empty space is a universal constant, the same for all observers.”

An inertial system is a coordinate system in which all isolated bodies, not acted on by any forces, move with a uniform velocity. Different inertial systems move relative to each other with some uniform velocity. So different inertial systems are like different observers moving relative to each other with constant velocity.

### 4.1.4 The relativity of simultaneity

Einstein’s postulates demand a completely new view of time. They show that two events which are simultaneous according to one observer may not be simultaneous according to another.

**Example:** A man with a flash-lamp stands in the centre of a railway carriage. The carriage is part of a train travelling along the track. At a given instant, the man makes the lamp emit a pair of light-pulses, one of which travels forward to the front of the carriage, while the other travels backward to the back of the carriage. The man with the lamp sees both pulses travel with the same speed $c$, and, since he is in the centre of the carriage, he sees the two pulses arrive simultaneously at the two ends of the carriage.

The same events are observed by another man standing beside the track. According to him, the light pulses do not arrive simultaneously at the two ends of the carriage. This is because the whole carriage is moving forward, so the backward-travelling pulse has less far to go than the forward-travelling pulse - the back of the train comes to meet it.

The arrivals of the light pulses at the front and back of the carriage are thus simultaneous for one observer (on the train) but not simultaneous for the other (beside the track).

The consequence of Einstein’s postulates illustrated by this example is called the “relativity of simultaneity”. It completely destroys Newton’s idea of absolute time flowing uniformly on.

### 4.2 Inertial systems and the Lorentz transformation

In order to understand the implications of Einstein’s postulates, we must examine the relationship between the descriptions of an event reported by observers who move relative to each other. To do this, we first define what is meant by an “inertial coordinate system”
and an “event”. Using these definitions we will summarise the “Galilean” transformation relating the specifications of an event according to the ideas before Einstein. Finally we derive the “Lorentz” transformation required to satisfy Einstein’s postulates.

4.2.1 Events and transformations

An event is something that happens at a particular point in space at a particular time. For example, the emission of a sharp pulse of light by a man with a lamp is an “event.” The arrival of the pulse at the end of a railway carriage is an event. The emission of an alpha-particle by a radioactive nucleus, and the decay of an unstable elementary particle like a muon or a pion are events.

In order to specify an event, we need to give its position in three-dimensional space, and its time. For present purposes, it is most convenient to use Cartesian coordinates, so that an event is specified by the four numbers $(x, y, z, t)$.

The same event can be observed by observers who are in motion relative to each other. That is, each observer has their own Cartesian coordinate system and their own clock. Then for a given event, the specifications of that event by two observers will be different: they give different numbers $(x, y, z, t)$ and $(x', y', z', t')$ to the event. Our aim is to give the relationship between these two specifications. This relationship is a linear transformation, like those we dealt with earlier. Thus there are matrices which give $(x', y', z', t')$ in terms of $(x, y, z, t)$, or vice versa. These will be $4 \times 4$ matrices, similar to the $3 \times 3$ matrices which describe rotations.

4.2.2 Inertial systems

In Newtonian mechanics, a special significance is given to “isolated” bodies, i.e. bodies that experience no forces because they do not interact with the rest of the world. Newton’s first law says that isolated bodies move with a uniform velocity. However, this statement needs to be made more precise, because velocity has to be relative to some observer. It is more complete and exact to state Newton’s first law in the form: “There are observers for whom all isolated bodies move with a uniform velocity.” Such observers are called inertial observers.

Here is the definition of an inertial coordinate system:

“An inertial coordinate system is a system of coordinates such that all isolated bodies move with uniform velocity in the coordinate system.”

In other words, an inertial coordinate system is a coordinate system in which an inertial observer is at rest. For brevity, an inertial coordinate system is usually just called an “inertial system.”
4.2.3 The Galilean transformation

Consider two inertial observers, moving relative to each other with speed \( v \). One observer, called \( S \), has a Cartesian coordinate system in which the Cartesian components of position are called \((x, y, z)\). The observer \( S \) has a clock fixed at the origin of the coordinate system, which measures time \( t \). The other observer, \( S' \), has a Cartesian coordinate system in which positions are denoted by \((x', y', z')\), and a clock fixed at the origin of the coordinate system, that measures time \( t' \). The length and time units used by the two observers are identical, as are their measuring sticks and clocks. The \( x \) and \( x' \) axes of the two systems are parallel to each other, and the same is true of the \( y \) and \( y' \) axes, and the \( z \) and \( z' \) axes. The origins of the two inertial systems coincide at \( t = 0 \), and the clocks of the two systems are synchronised so that at \( t = 0 \), \( t' = 0 \). The direction of motion of \( S' \) relative to \( S \) is along the positive \( x \)-axis.

There is an event, which, according to \( S \), is at \((x, y, z, t)\), and according to \( S' \) is at \((x', y', z', t')\). What is the formula relating \((x, y, z, t)\) to \((x', y', z', t')\)?

According to Galileo and Newton, at time \( t \), the origin of the inertial system \( S' \) is at the position \( x = vt \) as measured by \( S \). So at time \( t \), positions measured in \( S' \) are simply offset by this displacement \( x = vt \) along the \( x \)-direction. We therefore have:

\[
\begin{align*}
x' &= x - vt \\
y' &= y \\
z' &= z \\
t' &= t.
\end{align*}
\]

Equations 4.1 together constitute the “Galilean transformation.”. It can be represented as a \(4 \times 4\) matrix:

\[
\begin{pmatrix}
1 & 0 & 0 & -v \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

which would operate on a four-dimensional column vector \((x, y, z, t)\) such that

\[
\begin{pmatrix}
x' \\
y' \\
z' \\
t'
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & -v \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
x \\
y \\
z \\
t
\end{pmatrix}
\]

4.2.4 The Lorentz transformation

According to Einstein’s postulate, the Galilean transformation cannot be correct. Suppose a light pulse is emitted from the point \( x = y = z = 0 \) at time \( t = 0 \).
This event is also specified by
\[ x' = y' = z' = 0 \text{ at time } t' = 0; \]
that is the origins of the two frames \( S \) and \( S' \) coincide at \( t = t' = 0 \). Now consider the event in which the light-pulse arrives at the point \( x = x_0 \) on the \( x \)-axis of inertial system \( S \). Since observer \( S \) sees the light-pulse travel with speed \( c \), the time of arrival is \( t = x_0/c \).

So observer \( S \) specifies the arrival event as \( (x_0, 0, 0, x_0/c) \).

According to the Galilean transformation the same arrival event as specified by \( S' \) is \( (x_0 - vt, 0, 0, x_0/c) \).

This means that according to \( S' \) the speed at which the pulse travelled was
\[ (x_0 - vt)/(x_0/c) = c - v(ct/x_0). \]

But \( x_0 = ct \), so the speed of the pulse in \( S' \) is \( c - v \). This contradicts Einstein’s postulate.

It is possible to modify the Galilean transformation in such a way that Einstein’s postulate is true. However, there are strong limitations on the modifications that can be made:

- The equations expressing \((x', y', z', t')\) in terms of \((x, y, z, t)\) must remain linear. The reason is that the two coordinate systems are inertial. A body moving with uniform velocity in one system must also move with a uniform velocity in the other. This can only be so with linear relations.

- The symmetry properties of space must be respected. For example, an event that is observed to lie on the \( x \)-axis in \( S \) must also lie on the \( x \)-axis in \( S' \). Suppose that \( y = z = 0 \) and \( y' \neq 0 \) and/or \( z' \neq 0 \). Then the transformation would break the symmetry expected for rotations around the \( x \)-axis. Similarly, events on the \( x' \)-axis in \( S' \) must lie on the \( x \)-axis in \( S \). As another example of symmetry, suppose we have two events which in \( S \) are specified by \((x_0, y_0, z_0, t_0)\) and \((x_0, -y_0, z_0, t_0)\), so that they differ by reflection in the \( x-z \) plane. Then the two events must have the same value of \( x' \) in \( S' \); otherwise the transformation would break the symmetry expected for reflections in the \( x - z \) plane.

In light of this, consider the equation for \( x' \) in terms of \( x, y, z \) and \( t \). Since it is linear, we must have:
\[ x' = Ax + Bt + \alpha y + \beta z, \]
where \( A, B, \alpha \) and \( \beta \) are coefficients that may depend on the relative velocity \( v \). (There is no additive constant on the RHS, because we must have \( x' = 0 \) when \( x = y = z = t = 0 \).) By symmetry, we must have \( \alpha = \beta = 0 \); this follows from the requirement that events in \( S \) related by reflection in a plane containing the \( x \)-axis must have the same value of \( x' \).
For similar reasons, the equation for \( t' \) in terms of \( x, y, z \) and \( t \) must be:

\[
t' = Cx + Dt,
\]

with no terms in \( y \) and \( z \). Symmetry also dictates that we must have \( y' = y \) and \( z' = z \).

The most general form of the transformation allowed by symmetry is thus:

\[
\begin{align*}
x' &= Ax + Bt, \\
y' &= y, \\
z' &= z, \\
t' &= Cx + Dt.
\end{align*}
\]

or, in matrix form,

\[
\begin{pmatrix}
x' \\
y' \\
z' \\
t'
\end{pmatrix}
=
\begin{pmatrix}
A & 0 & 0 & B \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
C & 0 & 0 & D
\end{pmatrix}
\begin{pmatrix}
x \\
y \\
z \\
t
\end{pmatrix}.
\]

Expressions for \( A, B, C \) and \( D \) can be obtained in the following four steps:

1. The observer in \( S \) sees the origin of the coordinate system \( S' \) moving along the \( x \)-axis with speed \( v \). This means that if an event occurs at the origin of \( S' \) (for which \( x' = 0 \), by definition), then the \( x \) and \( t \) specifying this event must satisfy \( x/t = v \). From eq. 4.2, this requires that

\[
B = -Av.
\]

2. Similarly, the observer \( S' \) sees the origin of \( S \) moving with velocity \(-v\). This means that an event at the origin of \( S \) (for which \( x = 0 \), by definition) must have \( x'/t' = -v \). From eqs. 4.2 and 4.5, this means that \( B/D = -v \). Comparing this with eq. 4.7, we see that:

\[
D = A.
\]

3. A light pulse emitted at \( t = t' = 0 \) from the origin along the \( x \)-axis must be observed to travel at speed \( c \) in both \( S \) and \( S' \). Consider the event in which the light-pulse arrives at point \( x \) at time \( t \). In \( S' \), this same event is specified by \( x' \) and \( t' \). We must have both \( x/t = c \) and \( x'/t' = c \). Dividing eq. 4.2 by eq. 4.5, we have:

\[
\frac{x'}{t'} = c = \frac{Ax + Bt}{Cx + Dt} = \frac{Ac + B}{Cc + D}.
\]

Now we use eq. (4.7) and (4.8) to express \( B \) and \( D \) in terms of \( A \):

\[
c = \frac{c - v}{1 + cC/A}.
\]

Solving this for \( C/A \), we get:

\[
C = -\frac{v}{c^2} A.
\]
4. Finally, consider a light pulse emitted at \( t = t' = 0 \) from the origin along the \( y \) axis. This pulse must be observed to travel with speed \( c \) in both \( S' \) and \( S_0 \). Consider the event in which the light-pulse arrives at point \( y \) on the \( y \)-axis. Since \( x = 0 \), we have \( x' = Bt, \ y' = y, \ t' = Dt \). Now the distance travelled by the pulse in \( S_0 \) is \( \sqrt{(x')^2 + (y')^2} \), so that:

\[
\frac{(x')^2 + (y')^2}{(t')^2} = c^2 = \frac{B^2t^2 + y^2}{D^2t^2} = \left( \frac{B}{D} \right)^2 + \frac{c^2}{D^2},
\]

using the fact that \( y/t = c \). Now write \( B = -Av \) and \( D = A \) to obtain:

\[
c^2 = v^2 + \frac{c^2}{A^2}.
\]

Solving for \( A \), we obtain:

\[
A = \frac{1}{\sqrt{1 - v^2/c^2}}
\]

(Note that we must take the positive square, since for small \( v/c \) we must recover the Galilean transformation, so that \( A \to 1 \) as \( v/c \to 0 \).)

From the four eq.s 4.7, 4.8, 4.9 and 4.10 we have expressions for \( A, B, C \) and \( D \), which are:

\[
A = \frac{1}{\sqrt{1 - v^2/c^2}}, \quad B = -\frac{v}{\sqrt{1 - v^2/c^2}}
\]
\[
C = -\frac{v/c^2}{\sqrt{1 - v^2/c^2}}, \quad D = \frac{1}{\sqrt{1 - v^2/c^2}}
\]

Putting these expressions into eq.s 4.2, 4.3, 4.4 and 4.5 we have:

\[
x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}
\]
\[
y' = y
\]
\[
z' = z
\]
\[
t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}.
\]

These equations are known as the Lorentz transformations.

In relativistic physics, the quantity \( 1/\sqrt{1 - v^2/c^2} \) occurs very frequently, and it is usually called \( \gamma \) :

\[
\gamma \equiv \frac{1}{\sqrt{1 - v^2/c^2}}.
\]

Using \( \gamma \), we have:

\[
x' = \gamma(x - vt)
\]
\[
y' = y
\]
\[
z' = z
\]
\[
t' = \gamma(t - vx/c^2).
\]
and in terms of matrices

\[
\begin{pmatrix}
    x' \\
y' \\
z' \\
t'
\end{pmatrix} = \begin{pmatrix}
    \gamma & 0 & 0 & -\gamma v \\
    0 & 1 & 0 & 0 \\
    0 & 0 & 1 & 0 \\
    -\gamma v/c^2 & 0 & 0 & \gamma
\end{pmatrix}
\begin{pmatrix}
x \\
y \\
z \\
t
\end{pmatrix}.
\tag{4.13}
\]

Note that when the velocity \( v \) becomes very small compared with the velocity of light \( c \), the Lorentz transformation reduces to the Galilean transform, as we expect.

### 4.2.5 Invariance of the velocity of light

The Lorentz transformation formulae were derived above by considering light pulses directed along the \( x \)-axis and along the \( y \)-axis. But in fact the Lorentz transformation ensures that the velocity of light travelling in any direction is the same in all inertial frames.

Let a light pulse be emitted from the origin at \( t = 0 \). The pulse radiates in all directions. In the frame \( S \), at time \( t \), the pulse has reached all the points \((x, y, z)\) on the surface of a sphere of radius \( ct \); i.e all the points for which

\[x^2 + y^2 + z^2 - c^2t^2 = 0.\]

In the frame \( S' \), the same light-pulse should have reached all the points \(x', y', z'\) given by

\[(x')^2 + (y')^2 + (z')^2 - c^2(t')^2 = 0,
\]

where \((x', y', z', t')\) are related to \((x, y, z, t)\) by the Lorentz transformation.

From eq. 4.12, we have:

\[
(x')^2 + (y')^2 + (z')^2 - c^2(t')^2 = \gamma^2(x^2 - 2xvt + v^2t^2) + y^2 + z^2 - \gamma^2c^2(t^2 - 2vxt/c^2 + v^2x^2/c^4)
= \gamma^2(x^2 - c^2t^2 + c^2v^2t^2/c^2) + y^2 + z^2
= \gamma^2(x^2 - c^2t^2)(1 - v^2/c^2) + y^2 + z^2
= x^2 + y^2 + z^2 - c^2t^2.
\]

This demonstrates that the set of events specified by \((x, y, z, t)\) in \( S \) such that \(x^2 + y^2 + z^2 - c^2t^2 = 0\), is specified by \((x', y', z', t')\) in \( S' \) such that \((x')^2 + (y')^2 + (z')^2 - c^2(t')^2 = 0\). This is the same thing as saying that a light-pulse spreads out radially with speed \( c \) in both inertial frames, and hence in all inertial frames.

What we have in fact shown is that the quantity \(x^2 + y^2 + z^2 - c^2t^2\) is **invariant** under Lorentz transformations. Note that apart from the \( c \) and minus sign, this quantity actually looks a bit like the length of a four-dimensional coordinate vector \((x, y, z, t)\).
4.2.6 Four-vectors

The Lorentz transformations mix up space and time components in a manner analogous to the way in which rotations mix up space components. We have already written them as matrix transformations on a column vector.

"Four-vectors are objects which transform via the Lorentz matrices when boosted between inertial frames."

By analogy with normal vectors, four-vectors define a "direction" in space-time and have an invariant "length". Scalar products of four-vectors are invariant under Lorentz transformations. The only difference (apart from the fact that they are a four-dimensional not three-dimensional) is the minus sign which appears in front of the fourth term. This is taken into account (or explained by, if you prefer) by the fact that the metric (see section 3.9) of space-time is not just the identity matrix, but is given by:

\[
G = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\]

so that the scalar product of two four-vectors is:

\[
V \cdot W = V^T G W
\]

\[
= (V_1 V_2 V_3 V_4) \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\begin{pmatrix}
W_1 \\
W_2 \\
W_3 \\
W_4
\end{pmatrix}
\]

\[
= V_1 W_2 + V_2 W_2 + V_3 W_3 - V_4 W_4
\]

The space-time coordinates of a point are one important four-vector. In fact, to keep Lorentz transformations dimensionless, it is usual to write the coordinate four-vector as:

\[
\begin{pmatrix}
x \\
y \\
z \\
ct
\end{pmatrix}
\]

and the Lorentz transformations as

\[
\begin{pmatrix}
x' \\
y' \\
z' \\
ct'
\end{pmatrix}
= \begin{pmatrix}
\gamma & 0 & 0 & -\gamma v/c \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
-\gamma v/c & 0 & 0 & \gamma
\end{pmatrix}
\begin{pmatrix}
x \\
y \\
z \\
ct
\end{pmatrix}
\]

\[\text{(4.14)}\]

We will use this form of the transformation most often from now on. There are other four-vectors in physics - most importantly the energy-momentum four-vector, which we will meet later.
4.3 Consequences of the Lorentz transformation

4.3.1 Length contraction

Before Einstein, it was assumed without question that the length of an object is a fixed quantity, irrespective of how the object moves. According to the Lorentz transformation required by Einstein’s postulates, this is not correct. To see this, we have to define carefully what we mean by the length of a moving object.

Suppose we have a stick lying parallel to the $x$-axis, and suppose first that the stick is at rest. Then we note the $x$-coordinate of the left-hand end of the stick, call it $x_A$, and the $x$-coordinate of the right-hand end, call it $x_B$. The length of the stick $L_0$ is then the difference of $x_B$ and $x_A$:

$$L_0 = x_B - x_A.$$  

Since the stick is at rest, the coordinates $x_A$ and $x_B$ do not have to be measured at the same moment in time, because they do not change with time.

Now suppose the stick is moving to the right parallel to the $x$-axis, and hence parallel to its own length, with velocity $v$. At some instant in time, we measure $x_B$ and $x_A$, and define the length to be the difference of the two:

$$L = x_B - x_A.$$  

Since the stick is moving, it is clearly essential that the measurements of $x_A$ and $x_B$ are made at the same instant of time. If we did not insist on this, the measurement of the length of a moving object would not be well defined.

Now we ask: for a given stick, does the length $L$ defined by the above procedure depend on the velocity of the stick? And how is the length $L$ related to the length $L_0$ measured when the stick is at rest?

To answer this, let $S'$ be the inertial system in which the stick is at rest, with $S'$ moving along the positive $x$-axis with speed $v$ relative to inertial system $S$. Consider the two events in which the $x$-coordinates of the two ends of the stick are measured simultaneously in inertial system $S$. The four-vector coordinates of these two events in $S$ are

$$\begin{pmatrix} x_A \\ 0 \\ 0 \\ ct \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_B \\ 0 \\ 0 \\ ct \end{pmatrix}.$$  

According to our definition of the length of a moving object, the times $t$ of the two measurements are identical. These two measurement events, as specified in inertial system $S'$, are at

$$\begin{pmatrix} x'_A \\ 0 \\ 0 \\ ct'_A \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x'_B \\ 0 \\ 0 \\ ct'_B \end{pmatrix}.$$  

105
Since the measurement events are simultaneous in $S$, they are not simultaneous in $S’$. However, since the stick is at rest in $S’$, the measurements of $x’_A$ and $x’_B$ needed to determine the length do not have to be simultaneous in $S’$.

These four vectors are related via the Lorentz boost between the two frames (eq. 4.14):

$$
\begin{pmatrix}
  x'_{A} \\
  0 \\
  ct'_{A}
\end{pmatrix}
= 
\begin{pmatrix}
  \gamma & 0 & 0 & -\gamma v/c \\
  0 & 1 & 0 & 0 \\
  -\gamma v/c & 0 & 0 & \gamma
\end{pmatrix}
\begin{pmatrix}
  x_{A} \\
  0 \\
  ct
\end{pmatrix}
$$

and

$$
\begin{pmatrix}
  x'_{B} \\
  0 \\
  ct'_{B}
\end{pmatrix}
= 
\begin{pmatrix}
  \gamma & 0 & 0 & -\gamma v/c \\
  0 & 1 & 0 & 0 \\
  -\gamma v/c & 0 & 0 & \gamma
\end{pmatrix}
\begin{pmatrix}
  x_{B} \\
  0 \\
  ct
\end{pmatrix}
$$

So, looking at the first element:

$$
x'_{A} = \gamma (x_{A} - vt)
$$

$$
x'_{B} = \gamma (x_{B} - vt).
$$

Subtracting the two equations, we have:

$$
L_{0} = x'_{B} - x'_{A} = \gamma (x_{B} - x_{A}) = \gamma L.
$$

From the definition of $\gamma$, eq. 4.11, this gives:

$$
L = L_{0} \sqrt{1 - v^{2}/c^{2}}.
$$

(4.15)

The length $L$ of a moving object is therefore shorter than the length $L_{0}$ of this object when it is at rest by the factor $\sqrt{1 - v^{2}/c^{2}}$. This effect is called “relativistic length contraction”.

**Note:** Relativistic length contraction has nothing to do with forces acting on the object. It is not shortened by some force that is compressing it! The contraction occurs simply because if it did not occur the velocity of light in frames $S$ and $S’$ would not be the same.

The length $L_{0}$ of an object measured when it is at rest is called the “rest-length”.

### 4.3.2 Orientation of a moving stick

Length contraction occurs only along the direction of motion. A stick of rest-length $L_{0}$ oriented along the $y$-axis, still has exactly the same length $L_{0}$ if it moves in the $x$-direction. This means that a rectangular object moving along the $x$-axis with its edges parallel to the $x$- and $y$-axes, contracts only along the $x$-direction.

As an application of these ideas, consider a stick of rest-length $L_{0}$ at rest in inertial system $S’$. The stick is oriented at an angle $\theta_{0}$ to the $x$-axis, as measured in $S’$. What is the length of the stick and at what angle is it inclined to the $x$-axis, when observed from inertial system $S’$?
To answer this, let the positions of the two ends of the stick in system $S$ be at the positions $(x'_A = 0, y'_A = 0)$ and $(x'_B = L_0 \cos \theta_0, y'_B = L_0 \sin \theta_0)$. Let the positions of the two ends of the stick be measured simultaneously at the instant $t = 0$ in $S$. Then from the Lorentz transformation, eq.(4.14), we have:

\[
\begin{pmatrix}
0 \\
0 \\
0 \\
c t'_A
\end{pmatrix} = \begin{pmatrix}
\gamma & 0 & 0 & -\gamma v/c \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
-\gamma v/c & 0 & 0 & \gamma
\end{pmatrix} \begin{pmatrix}
x_A \\
y_A \\
y_A \\
0
\end{pmatrix}
\]  

(4.16)

and

\[
\begin{pmatrix}
L_0 \cos \theta_0 \\
L_0 \sin \theta_0 \\
0 \\
c t'_B
\end{pmatrix} = \begin{pmatrix}
\gamma & 0 & 0 & -\gamma v/c \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
-\gamma v/c & 0 & 0 & \gamma
\end{pmatrix} \begin{pmatrix}
x_B \\
y_B \\
y_B \\
0
\end{pmatrix}
\]  

(4.17)

From eq.(4.16) we get $x_A = 0$, $y_A = 0$, and from eq.(4.17) we get

\[
\begin{align*}
x'_B &= L_0 \cos \theta_0 = \gamma x_B \\
y'_B &= L_0 \sin \theta_0 = y_B.
\end{align*}
\]

From this the length of the stick measured in $S$ is:

\[
L = (x'_B^2 + y'_B^2)^{\frac{1}{2}} = \left[ L_0^2 \cos^2 \theta_0 \left( 1 - \frac{v^2}{c^2} \right) + L_0^2 \sin^2 \theta_0 \right]^{\frac{1}{2}}
\]

\[
= L_0 \left[ 1 - \frac{v^2}{c^2} \cos^2 \theta_0 \right]^{\frac{1}{2}}.
\]

The angle $\theta$ at which the stick is inclined to the $x$-axis in $S$ is given by:

\[
\begin{align*}
\theta &= \tan^{-1}(y_B/x_B) \\
&= \tan^{-1}(\gamma \tan \theta_0).
\end{align*}
\]

So the moving rod is both contracted and rotated.

### 4.3.3 Time dilation

Now consider the rate at which a moving clock runs. A clock is at the origin in inertial system $S'$, which, as usual, moves with velocity $v$ along the positive $x$-axis relative to inertial system $S$. The moving clock records the lapse of a certain time $\Delta t'$, and we want to know the corresponding time lapse recorded in $S$. To put it more concretely, suppose the moving clock ticks every second. What is the time interval between the ticks according to an observer at rest in $S$?

To answer this question, it is helpful to have the Lorentz equations expressing $x$, $y$, $z$ and $t$ in terms of $x'$, $y'$, $z'$ and $t'$, i.e. the “inverse” equations. These inverse equations can be obtained several ways: for example, by finding the inverse matrix to the Lorentz matrix, by treating the Lorentz equations as simultaneous equations and solving them, or by noting
that $S$ moves with velocity $-v$ relative to $S'$, so we can get the inverse equations simply by swapping $x$ and $x'$, $y$ and $y'$, $z$ and $z'$, and $t$ and $t'$, while changing the sign of $v$. We therefore have:

$$
\begin{pmatrix}
  x \\
  y \\
  z \\
  ct
\end{pmatrix} =
\begin{pmatrix}
  \gamma & 0 & 0 & \gamma v/c \\
  0 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 \\
  \gamma v/c & 0 & 0 & \gamma
\end{pmatrix}
\begin{pmatrix}
  x' \\
  y' \\
  z' \\
  ct'
\end{pmatrix}.
$$

(4.18)

Now consider two ticks of the clock which occur at the instants $t^\prime = t^\prime_A$ and $t^\prime_B$ as recorded in frame $S'$. Since the clock is at the origin in $S'$, the two ticks constitute two events specified in $S'$ by

$$
\begin{pmatrix}
  0 \\
  0 \\
  ct^\prime_A
\end{pmatrix}
\text{ and }
\begin{pmatrix}
  0 \\
  0 \\
  ct^\prime_B
\end{pmatrix}
$$

The corresponding events as specified in $S$ are at times $t_A$ and $t_B$ given, according to eq.s 4.18, by:

$$
ct_A = \gamma ct^\prime_A , \quad ct_B = \gamma ct^\prime_B.
$$

So the two ticks of the clock separated by the time difference $\Delta t^\prime = t^\prime_B - t^\prime_A$ in $S'$ are separated by time difference $\Delta t$ in $S$ given by:

$$
\Delta t = t_B - t_A = \gamma \Delta t^\prime.
$$

Since $\gamma = 1/\sqrt{1 - v^2/c^2}$ is greater than unity, this means that $\Delta t > \Delta t^\prime$. So, for example, if the clock ticks once a second, then when it is moving with velocity $v$ a stationary observer measures a time greater than 1 second between ticks.

One experimental example of time dilation comes from the decay of muons produced by cosmic rays in the upper atmosphere. The muon is an unstable analogue of the electron, having a lifetime of $2.2\ \mu s$ when measured in an inertial system in which it is at rest. Travelling at the speed of light $c = 3 \times 10^8$ m s$^{-1}$ we might expect them to travel a maximum distance of only 600 m before decaying. However, in fact they succeed in travelling several km from the upper atmosphere to the Earth’s surface. This discrepancy is explained by time dilation. See e.g. Halliday & Resnick (“Fundamentals of Physics”) for details.

### 4.3.4 The relativity of simultaneity

We already saw that events that are simultaneous in one inertial system may not be simultaneous in another (the man with a flash-lamp in the railway carriage). We can return to the question of simultaneity now in terms of the Lorentz transformation.

Suppose we have two events, which in system $S$ are specified by

$$
\begin{pmatrix}
  x_A \\
  y_A \\
  z_A \\
  ct_A
\end{pmatrix}
\text{ and }
\begin{pmatrix}
  x_B \\
  y_B \\
  z_B \\
  ct_B
\end{pmatrix}
$$

108
and in system $S'$ are specified by
\[
\begin{pmatrix}
  x'_A \\
  y'_A \\
  z'_A \\
  c t'_A
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
  x'_B \\
  y'_B \\
  z'_B \\
  c t'_B
\end{pmatrix}
\]

According to the inverse Lorentz transformation, eq.s (4.18), the time interval $t_B - t_A$ between the events in $S$ is given by:
\[
c(t_B - t_A) = \gamma ((c t'_B - c t'_A) + (x'_B - x'_A)v/c).
\]

Now suppose the events are simultaneous in system $S$, so that $t_B - t_A = 0$. Then in system $S'$ the time interval is:
\[
t'_B - t'_A = -(x'_B - x'_A)v/c^2.
\]

This means that the events are simultaneous in $S'$ only if $x'_A = x'_B$. Otherwise, the order in which the events occur depends on the sign of $x'_B - x'_A$ and on the sign of the relative velocity $v$. 

109
4.3.5 The interval between events

The time interval between two events and the spatial distance between events depend on the inertial system in which the events are observed. Nevertheless, four-vectors provide a way of characterising the interval between two events that does not depend on the inertial system.

Consider two coordinate four-vectors and the difference between them:

\[
\begin{pmatrix}
  x_A \\
  y_A \\
  z_A \\
  ct_A \\
\end{pmatrix} - \begin{pmatrix}
  x_B \\
  y_B \\
  z_B \\
  ct_B \\
\end{pmatrix} = \begin{pmatrix}
  x_A - x_B \\
  y_A - y_B \\
  z_A - z_B \\
  ct_A - ct_B \\
\end{pmatrix} = q
\]

Since the Lorentz transformation is linear, the difference of two four-vectors is also a four-vector. Thus the squared magnitude of \( q \):

\[
q^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2 \Delta t^2,
\]

where \( \Delta x = x_A - x_B, \Delta y = y_A - y_B, \Delta z = z_A - z_B \) and \( \Delta t = t_A - t_B \), is invariant. Thus for two given events, the quantity \( q^2 \) is the same in all inertial frames.

Note that \( q^2 \) can be positive, negative or zero:

- \( q^2 = 0 \) : then the distance between the events \( \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \) is equal to \( c|\Delta t| \), so that a light signal can pass from one event to the other.

- \( q^2 < 0 \) : in this case, \( \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \) is less than \( c|\Delta t| \), so that a particle can travel from one event to the other at a speed less than the velocity of light. Put another way, there is an inertial system in which \( \Delta x = \Delta y = \Delta z = 0 \), so that both events occur at the same position.

- \( q^2 > 0 \) : in this case, \( \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \) is greater than \( c|\Delta t| \), so it is impossible for a signal to pass from one event to the other at less than the velocity of light. In this case, there is an inertial system in which the two events are simultaneous.

If \( q^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2 \Delta t^2 < 0 \), the interval between the two events is called “time-like”. The events always occur in the same time order in every inertial frame.

If \( q^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2 \Delta t^2 > 0 \), the interval between the two events is called “space-like”. The time order in which the events occur depends on inertial system, and there is an inertial system in which the events are simultaneous.

Note that for anything travelling at the speed of light, the interval between all points on its path is zero.
4.3.6 Addition of velocities

Suppose an object is moving with a certain speed in a certain direction as observed from inertial system $S$, what is its speed and direction as observed from $S'$? As usual, the Cartesian axes of $S'$ are parallel to the Cartesian axes of $S$, and $S'$ moves at speed $v$ along the positive $x$-direction relative to $S$.

To describe speed and direction, we need to consider the velocity vector. So in system $S$, let the velocity vector of the moving body be $u = (u_x, u_y, u_z)$. In system $S'$, the velocity vector of the moving body is $u' = (u'_x, u'_y, u'_z)$. Our task is to find formulae expressing $u'_x$, $u'_y$ and $u'_z$ in terms of $u_x$, $u_y$ and $u_z$.

We can derive the required formulae from the Lorentz transformation by thinking in terms of events. Suppose that, as observed from $S$, at time $t_A$, the body is at $(x_A, y_A, z_A)$, and at time $t_B$ it is at $(x_B, y_B, z_B)$. These are two events. The velocity with which the body moves in $S$ has Cartesian components given by:

$$
u_x = \frac{x_B - x_A}{t_B - t_A}, \quad u_y = \frac{y_B - y_A}{t_B - t_A}, \quad u_z = \frac{z_B - z_A}{t_B - t_A}. \quad (4.19)$$

Now let these same events $A$ and $B$ be specified in system $S'$ by $(x'_A, y'_A, z'_A, t'_A)$ and $(x'_B, y'_B, z'_B, t'_B)$. Then the velocity vector in $S'$ is given by:

$$u'_x = \frac{x'_B - x'_A}{t'_B - t'_A}, \quad u'_y = \frac{y'_B - y'_A}{t'_B - t'_A}, \quad u'_z = \frac{z'_B - z'_A}{t'_B - t'_A}. \quad (4.20)$$

Now use the Lorentz transformation (eq. 4.14) to write primed quantities in terms of unprimed quantities. To relate $u'_x$ to $u_x$, note that:

$$x'_B - x'_A = \gamma (x_B - x_A - v(t_B - t_A)), \quad t'_B - t'_A = \gamma \left(t_B - t_A - \frac{v}{c^2}(x_B - x_A)\right),$$

where, as usual, $\gamma = 1/\sqrt{1 - v^2/c^2}$. Dividing $x'_B - x'_A$ by $t'_B - t'_A$, we get:

$$u'_x = \frac{x'_B - x'_A}{t'_B - t'_A} = \frac{x_B - x_A - v(t_B - t_A)}{t_B - t_A - (v/c^2)(x_B - x_A)} \quad (4.21)$$

Now divide top and bottom by $t_B - t_A$:

$$u'_x = \frac{u_x - v}{1 - u_x v/c^2}. \quad (4.22)$$

Doing the same thing for $u'_y$:

$$u'_y = \frac{y'_B - y'_A}{t'_B - t'_A} = \frac{y_B - y_A}{\gamma(t_B - t_A - (v/c^2)(x_B - x_A))}$$

$$= \frac{1}{\gamma} \cdot \frac{u_y}{1 - u_x v/c^2}. \quad (4.23)$$
Similarly, for $u_z'$:

$$u_z' = \frac{1}{\gamma} \cdot \frac{u_z}{1 - u_x v/c^2}.$$  \hspace{2cm} (4.24)

The corresponding expressions for $(u_x', u_y', u_z')$ in terms of $(u_x', u_y', u_z')$ can be found by swapping $u_x$ and $u_x'$, $u_y$ and $u_y'$ and $u_z$ and $u_z'$, and changing the sign of $v$:

$$u_x = \frac{u_x' + v}{1 + u_x' v/c^2}$$ \hspace{2cm} (4.25)

$$u_y = \frac{1}{\gamma} \cdot \frac{u_y'}{1 + u_y' v/c^2}$$ \hspace{2cm} (4.26)

$$u_z = \frac{1}{\gamma} \cdot \frac{u_z'}{1 + u_z' v/c^2}.$$ \hspace{2cm} (4.27)

These formulae can also be derived by the multiplying of the two Lorentz matrices; see the problem sheets.

Note that the three components of velocity do not transform like the spatial components of a four-vector.

### 4.3.7 Examples of addition of velocities:

Suppose a spaceship travels directly away from the Earth with a speed $v = 0.5c$. It fires a space-probe directly ahead of it (i.e. away from the Earth). The speed of the space-probe is $u = 0.5c$ as seen from the spaceship. What is the speed of the space-probe as seen from the Earth? If we used common sense, instead of relativity, we would just add the velocities, and conclude that the space-probe moves at the speed of light as seen from the Earth. But this is incorrect!

To apply the formulae for relativistic addition of velocities, let system $S$ be fixed on the Earth, system $S'$ be fixed on the spaceship. Let the $x$-axis be in the direction in which the spaceship travels. Then $v$ is the speed of the spaceship relative to the Earth, $u_x'$ is the speed of the space-probe as seen from the spaceship, and $u_x$ is the speed of the space-probe as seen from Earth. With $v = 0.5c$, $u_x' = 0.5c$, using eq. 4.25 we have:

$$u_x = \frac{c}{1 + \frac{v}{c}} = 0.8c.$$  

If instead of firing a space-probe with speed $u = 0.5c$, the spaceship had fired a pulse of light (obviously in this case $u' = c$), the speed of the pulse of light relative to Earth would be:

$$u_x = \frac{1.5c}{1 + 0.5} = c.$$  

This is what we expect, and simply confirms that the speed of light is the same for all observers.
4.3.8 Relativistic Doppler effect

The "Doppler effect" is the effect in which the apparent frequency of a wave-motion is altered by the motion of the source, or the observer, or both. It is familiar from everyday life. The pitch of the sound from a moving vehicle, e.g. the siren of a fire-engine, is raised when it approaches you, and is lowered when it recedes from you. The same effect happens with light sources. I derive here the relativistic formula for the shifts of frequency and wavelength of a monochromatic light source.

Consider a light source that emits a regular sequence of light pulses. In the rest-frame of the source, let the time interval between pulses be $T_0$. The light source now moves away from the observer along the observer’s line of sight with speed $v$. I will obtain a formula for the time interval $T_1$ between the arrival of consecutive pulses at the observer.

Note first that the sequence of light pulses is like a clock. Because of time dilation (section 4.3.3), the clock appears to run slow in the observer’s inertial frame, so that the time interval $T$ between the emission of consecutive pulses in the observer’s frame is:

$$T = \gamma T_0. \tag{4.28}$$

This interval $T$ is, of course, not the time interval $T_1$ between the arrival of pulses at the observer – it is the time interval between the emission of pulses. To obtain $T_1$, consider two consecutive pulses. The first pulse is emitted when the source is at some distance $d$ from the observer, so it takes a time $d/c$ to arrive. The second pulse is emitted at time $T$ later, by which time the distance from the source to the observer has increased to $d + vT$ (note, $d$, $v$ and $T$ are all as seen in the observer’s frame). So the time interval between the arrival of these two pulses at the observer is:

$$T_1 = T + \frac{d + vT}{c} - \frac{d}{c} = T \left(1 + \frac{v}{c}\right).$$

Noting that $\gamma = 1/\sqrt{(1 - v^2/c^2)} = 1/\sqrt{(1 - v/c)(1 + v/c)}$, we can also write this as:

$$T_1 = T_0 \sqrt{\frac{1 + v/c}{1 - v/c}}. \tag{4.29}$$

Now it is easy to go from this formula for the interval between the arrival of pulses to a formula for the Doppler shift of the frequency of a monochromatic light source.

A monochromatic source gives out a sinusoidal wave, and in the rest-frame of the source there is a time interval $T_0$ between the emission of consecutive crests of the wave. The frequency of the source in the rest-frame is $f_0 = 1/T_0$. To calculate the frequency $f$ of light received by the observer, we say $f = 1/T_1$, where $T_1$ is the time interval between the arrival of consecutive wave crests at the observer. But the formula relating $T_1$ and $T_0$ for wave crests must the same as the formula relating $T_1$ and $T_0$ for light pulses. So from eq. (4.29), the frequency $f$ seen by the observer is given by:

$$f = \frac{1}{T_1} = \frac{1}{T_0} \sqrt{\frac{1 - v/c}{1 + v/c}}.$$
Hence:

\[ f = f_0 \sqrt{\frac{1 - v/c}{1 + v/c}} \quad (4.30) \]

**Please note:** The signs of \( v \) appearing in this formula depend on whether the source is approaching or receding from the observer. In the version written above, \( v \) is the speed with which the source *recedes*. Some books (e.g. Kleppner and Kolenkow) give a formula for \( f \) in terms of \( f_0 \) with opposite signs of \( v \), but this is because they assume that the source is *approaching*.

The formula for \( f \) in terms of \( f_0 \) is easily converted to a formula for wavelengths. Suppose \( \lambda_0 \) is the wavelength of the light source as observed in its rest frame. Then \( \lambda_0 = c/f_0 \). If \( \lambda \) is the wavelength seen by the observer, then \( \lambda = c/f \), so for a source receding with speed \( v \), we have:

\[ \lambda = \lambda_0 \sqrt{\frac{1 + v/c}{1 - v/c}} \quad (4.31) \]

For a receding source, the frequency is lowered, and the wavelength is lengthened: we get a *red shift*. For an approaching source, frequency is raised, wavelength is shortened: a *blue shift*.

What happens for a source moving perpendicular to the line between it and the observer? In this case, unlike the non-relativistic case, there is still a *frequency shift* due to the time dilation in eq. 4.28. This shift is

\[ f = f_0/\gamma. \]
4.4 Mass, Energy and $E = mc^2$

Momentum and energy are absolutely central to physics, because they are conserved quantities. The total energy and the momentum of an isolated system are constant, and do not change with time. In modern physics, conservation laws are a cornerstone of fundamental theories, because they are intimately linked to symmetries.

The conservation of energy and momentum is maintained and generalised in special relativity. However, since in relativity we have seen that the addition of velocities is modified, it is reasonable to assume that there will be implications for the formulae for momentum and kinetic energy.

4.4.1 Why mass must depend on speed

Momentum is a vector. In Newtonian physics, the momentum of a particle is its mass $m$ times the velocity vector $v$:

$$p = mv.$$  

According to Newton’s laws of motion, the three Cartesian components of the total momentum of an isolated system of interacting particles are all conserved. For example, if two particles collide, the total momentum vector before the collision is equal to the total momentum after the collision. This is exactly true in the rest frame of the isolated system and by the principles of special relativity it must therefore be true in any inertial frame, regardless of the relative velocity of the frame with respect to any observer. We will use this fact to see how mass must depend on speed.

Consider the collision of two particles which are identical to each other. The first figure shows the collision as observed in the inertial frame in which the centre of mass of the two particles is at rest.

![Collision in centre-of-mass inertial frame.](image)

The collision is elastic, which means that it is reversible in time. If, after the collision, the velocities of the particles are reversed, exactly the same collision will occur in reverse. Take the $x$-axis as being horizontal in the diagram, i.e. the line of symmetry running through the centre of mass, such that the trajectory of $A$ is the mirror image of the trajectory of $B$ (but in reverse time-order) when we reflect in the $x$-axis. The $y$-axis is the symmetry axis at right-angles passing through the centre of mass.

By conservation of momentum, the $x$-components of momentum of the two particles must be equal and opposite both before and after the collision, as must the $y$-components.
By symmetry, the $x$-component of velocity of each particle before the collision is equal to its $x$-component after the collision. This has to be true because (a) the particles are identical, and there is no reason for one particle to speed up while the other slows down and (b) because the collision is reversible (symmetric in time) so we cannot have both particles going either faster or slower than they were before the collision. By similar arguments, the $y$-components of velocity of each particle before and after the collision are equal in magnitude, but have opposite signs.

Now go to the inertial frame which moves along the positive $x$-axis at a speed such that in this frame the $x$-component of velocity of $A$ before and after the collision is zero. Call this frame $S$ (see second figure).

In the $S$-frame, let the $x$-component of velocity of particle $B$ before and after the collision be called $-v$. Before and after the collision, particle $A$ is moving parallel to the $y$-axis with $y$-component of velocity that I call $-u_0$ and $u_0$ (equal and opposite, because they are equal and opposite in the centre-of-mass frame). The $y$-components of velocity of $B$ before and after the collision are called $u_1$ and $-u_1$.

Using the formula for velocities under Lorentz transformations, and the symmetries of the problem, we can get a formula for $u_1$ in terms of $u_0$ as follows: Consider only particle $B$, and look at it in the $S'$-frame (see the third picture):

that is, the inertial frame moving along the $x$-axis to the left relative to the $S$-frame at speed $v$. In the $S'$-frame, the $x$-component of velocity of $B$ before and after the collision is zero ($u'_x = 0$). Furthermore, the $y$-component of the velocity of $B$ in the $S'$-frame must be $u_0$ before and $-u_0$ after the collision, i.e. equal magnitude but opposite direction to the
$y$-component of the velocity of $A$ in the $S'$-frame. This must be true by symmetry. Hence, $u_1$ is the $y$-component of velocity in the $S'$-frame (i.e. $u_y = u_1$) of a particle which, in the $S'$ frame, is moving along the $y$-axis with velocity $u_y' = u_0$. Now use the relativistic formula relating $y$-component of velocities in different frames:

$$u_y = \frac{1}{\gamma} \frac{u_y'}{1 + u_x' v/c^2}.$$  

In the present case, $u_x'$ represents the $x$-component of velocity of particle $B$ in the $S'$-frame, which is zero, so we have:

$$u_y = u_1 = \frac{u_0}{\gamma} = u_0 \sqrt{1 - v^2/c^2}. \tag{4.32}$$

The crucial point to appreciate here is that $u_0$ and $u_1$ are not the same. The $y$-components of the velocities of $A$ and $B$ are equal in the centre-of-mass frame, but they are not the same in the $S$-frame or the $S'$-frame, because of the relativistic relations between velocities in different inertial frames.

If $u_1 \neq u_0$, this poses an extraordinary problem, if we believe that momentum is conserved. The problem can only be resolved by postulating that the mass of a particle depends on its speed. Here is the problem: the $y$-component of total momentum before the collision is $-m_A u_0 + m_B u_1$, and after the collision is $m_A u_0 - m_B u_1$, so that the momentum changes sign in the collision. Momentum can only be conserved if $m_A u_0 - m_B u_1 = 0$. But since $u_1 \neq u_0$, we must have $m_A \neq m_B$, even though the particles are identical in every way.

Faced with this problem, Einstein decided that conservation of momentum is such a fundamental principle that it must be maintained at all costs. In order to maintain it, he had to say that mass depends on speed.

### 4.4.2 Dependence of mass on speed

Now we derive a formula expressing how mass depends on speed. Note that we assume that mass depends only on the magnitude of velocity, and not on the individual Cartesian components of the velocity vector. (This must be true if we respect the principle of relativity that there are no preferred inertial frames, and hence no preferred directions).

Work in frame $S$. Since the speed of $A$ before and after the collision is $u_0$, we write $m_A = m(u_0)$. The speed of $B$ before and after the collision, denoted by $w$, is $w = \sqrt{u_1^2 + v^2}$, so we write $m_B = m(w)$. Then the requirement $m_A u_0 - m_B u_1 = 0$ can be expressed as:

$$u_0 m(u_0) = u_1 m(w). \tag{4.33}$$

Putting the formula from eq.(4.32) into eq.(4.33), we have:

$$m(u_0) = m(w) \sqrt{1 - v^2/c^2}, \tag{4.34}$$

where I have cancelled a factor $u_0$ on both sides.

We can now deduce from eq.(4.34) how $m$ must depend on speed. To do this, we consider what happens when $u_0$ becomes very small. On the left of eq.(4.34), we just get $m_0$, the
mass of the particle when it is at rest. On the right, we note that \( w \rightarrow v \) when \( u_0 \rightarrow 0 \). This gives:

\[
m(v) = \frac{m_0}{\sqrt{1 - v^2/c^2}} = \gamma m_0. \tag{4.35}
\]

[You should ask: with \( m(v) \) given by eq.(4.35), is eq.(4.34) exactly satisfied even when \( u_0 \) is not small? The answer is yes: you might like to try verifying this.]

Eq.(4.35) shows that the mass of a particle increases with speed, and (if \( m_0 \neq 0 \)) tends to infinity as the speed approaches the speed of light.

4.4.3 Total energy of a moving body

In Newtonian physics, a force \( F \) does work on a body, causes it to accelerate or decelerate, and changes its kinetic energy. In one dimension, if the body moves through distance \( \delta x \), the work done is \( F\delta x \), and this is equal to the change of kinetic energy \( \delta K \). If the body is moving with speed \( v \), then \( \delta x = v\delta t \), so the rate of change of kinetic energy is \( dK/dt = Fdx/dt = Fv \). But the force is the rate of change of momentum: \( F = dp/dt \). The momentum is \( p = mv \). In Newtonian physics, \( m \) is constant, so that \( dp/dt = mdv/dt \). Putting these facts together, we have:

\[
\frac{dK}{dt} = v \frac{dp}{dt} = mv \frac{dv}{dt} = \frac{d}{dt} \left( \frac{1}{2}mv^2 \right) . \tag{4.36}
\]

This shows that the kinetic energy is equal to \( \frac{1}{2}mv^2 \), if we require that \( K = 0 \) when \( v = 0 \).

Now we assume that we can apply the same argument in the context of relativity. It is still true that

\[
\frac{dK}{dt} = v \frac{dp}{dt} = v \frac{d}{dt}(mv) , \tag{4.37}
\]

but in order to evaluate \( d(mv)/dt \) we have to remember that \( m \) depends on \( v \). Using the formula \( m(v) = m_0/\sqrt{1 - v^2/c^2} \) (eq. 4.35), we have:

\[
\frac{d(mv)}{dt} = \frac{d}{dt} \frac{m_0v}{\sqrt{1 - v^2/c^2}} = m_0 \left[ \left( \frac{1 - v^2}{c^2} \right)^{-1/2} + \frac{v^2}{c^2} \left( \frac{1 - v^2}{c^2} \right)^{-3/2} \right] \frac{dv}{dt} = m_0 \left( \frac{1 - v^2}{c^2} \right)^{-3/2} \frac{dv}{dt} .
\]

Hence, from eq. (4.37):

\[
\frac{dK}{dt} = m_0v \left( 1 - \frac{v^2}{c^2} \right)^{-3/2} \frac{dv}{dt} = \frac{d}{dt} \frac{m_0v^2}{\sqrt{1 - v^2/c^2}} = \frac{d}{dt} \left( mc^2 \right) . \tag{4.38}
\]
Now we integrate this equation to give:

\[ K = mc^2 + A, \]

where \( A \) is an arbitrary constant. As in the Newtonian case, \( A \) is determined by requiring that \( K = 0 \) when \( v = 0 \). But when \( v = 0, m = m_0 \), the rest mass, so that:

\[ K = mc^2 - m_0c^2 \]  \hspace{1cm} (4.39)

This looks totally different from the Newtonian expression \( K = \frac{1}{2}mv^2 \), but in fact reproduces this formula if \( v/c \) is very small:

\[
K = mc^2 - m_0c^2 = m_0c^2 \left(1 - \frac{v^2}{c^2}\right)^{-1/2} - m_0c^2 \\
= \frac{1}{2}m_0v^2 \left(1 + \text{terms of order } (v/c)^2\right).
\]

Eq.(4.39) can be rewritten as:

\[ mc^2 = m_0c^2 + K. \]  \hspace{1cm} (4.40)

Einstein proposed a radical interpretation of eq. 4.38. He said the total energy \( E \) in a body is \( mc^2 \). This means that even when a body is stationary it contains a “rest energy” \( m_0c^2 \). When the body moves, its energy \( E = mc^2 \) increases because its mass \( m \) increases. The increase of energy is the kinetic energy, represented by the difference \( mc^2 - m_0c^2 \).

This is the origin of the equation

\[ E = mc^2. \]

This formula represents the total energy of a body, both the kinetic energy and an intrinsic energy that exists simply because the body has a mass. Put another way, we can say that energy and mass are the same thing, apart from a factor of \( c^2 \) (which is just a matter of the units used to specify \( E \) and \( m \)):

“Whatever has mass has energy, and whatever has energy has mass.”

### 4.4.4 Internal energy and mass

The implications of \( E = mc^2 \) are very remarkable. The equation implies that any form of energy at all is associated with mass. For example, if you compress a spring (e.g. when you wind up a clock), energy is stored in the spring. According to \( E = mc^2 \), a compressed spring has a larger rest-mass than a relaxed spring. If a compressed spring and a relaxed spring move at the same speed \( v \), the compressed spring has a greater momentum \( p = mv \). When you heat up an object, its mass increases. When you charge up a battery, you increase its mass, not because you add material to it (which you don’t) but because you add energy to it. If hydrogen and oxygen combine to form water, the number of protons, neutrons and electrons stays exactly the same, but the total mass decreases, because heat was given off. If two deuterons combine to form a helium nucleus, the number of protons and neutrons stays the same, but the total mass decreases.
Now let’s be sceptical! Suppose you don’t believe that a compressed spring has a bigger mass! I will convince you by discussing another collision problem.

This time, we have an inelastic collision between two identical bodies $A$ and $B$, each having rest mass $m_0$. In the inertial frame in which the centre of mass is stationary, they approach each other with equal and opposite velocities $v$ and $-v$ along the $x$-axis. When they collide, they stick together, and the combined body $AB$ is stationary, by symmetry. Let’s suppose that the kinetic energy lost goes into compressing springs inside the bodies.

According to Einstein, the total energy before the collision is

$$E = 2mc^2 = \frac{2m_0c^2}{\sqrt{1-v^2/c^2}}.$$ 

Since energy is conserved, Einstein says that this $E$ must be equal to $m_{AB}^0c^2$, where $m_{AB}^0$ is the mass of the combined object. This requires that $m_{AB}^0$ is not equal to $2m_0$; instead it is equal to

$$m_{AB}^0 = \frac{2m_0}{\sqrt{1-v^2/c^2}}.$$ 

Einstein claims that it is greater than $2m_0$ because the energy stored in the springs has a mass.

Now we’ll test this claim by considering the collision in the inertial frame in which body $A$ is initially at rest. In this frame, the initial velocity of body $B$ is

$$u = \frac{2v}{1+v^2/c^2}; \quad (4.41)$$

where we have used the usual formula for relativistic addition of velocities. The final velocity of the combined body $AB$ in this frame is $v$.

From this, the initial momentum, as seen in this inertial frame, is:

$$um(u) = \frac{m_0}{\sqrt{1-u^2/c^2}} \cdot \frac{2v}{1+v^2/c^2}.$$ 

The momentum after is:

$$\frac{m_{AB}^0}{\sqrt{1-v^2/c^2}}v,$$

where $m_{AB}^0$, as before, is the rest-energy of the combined body. Momentum conservation therefore demands that:

$$\frac{m_{AB}^0}{\sqrt{1-v^2/c^2}} = \frac{2m_0}{\sqrt{1-u^2/c^2}(1+v^2/c^2)}.$$ 

If we now insert the formula for $u$ (eq.4.41) into the right-hand side, after a little algebra, we obtain:

$$m_{AB}^0 = \frac{2m_0}{\sqrt{1-v^2/c^2}}.$$
This means that Einstein must be right! The rest mass of the combined particle cannot be $2m_0$. If it was, momentum would not be conserved.

This confirms that a compressed spring must have a greater rest mass than a relaxed spring, and the contribution of the stored energy to the mass is correctly given by $E = mc^2$. But if the energy stored in a spring has mass, then all other forms of energy must have a mass too. In the collision problem just considered, we could just as well have supposed that the kinetic energy lost goes into heating the bodies, charging up batteries, breaking up water into hydrogen and oxygen or breaking alpha-particles into deuterons.

### 4.4.5 Relativistic Energies

Relativistic effects are most important when particles travel at speeds comparable with the speed of light, or when their interaction energies are comparable with their rest energies. The rest energy $m_0c^2$ of an electron is approximately 0.51 MeV (1 MeV = 1 million electron volts). This is enormously greater than the energies that characterise chemical reactions (usually a few eV). The energy region of 1 MeV and upwards is the domain of nuclear and particle physics, and is precisely the region where the total energy of the particle can be much bigger than its rest energy. I will illustrate the application of relativistic ideas using some worked examples from nuclear and particle physics. Other areas of physics where such high energies are important include accelerator physics and cosmology.

As an illustration of the very large interaction energies that play a role in nuclear physics, consider the “fusion” reaction in which two deuterons combine to form a helium-4 nucleus:

$$2^2\text{H} \rightarrow ^4\text{He}.$$  

The deuteron $^2\text{H}$, a heavy isotope of hydrogen, consists of a proton and a neutron bound together by the strong nuclear force. The helium-4 nucleus $^4\text{He}$ consists of two protons and two neutrons bound together. The rest-mass of $^4\text{He}$ is 4.0026 u. (The atomic-mass unit u is defined so that the mass of the $^{12}\text{C}$ atom is exactly 12 u.) The mass of the deuteron $^2\text{H}$ is 2.0141 u. You can see from that the mass of $^4\text{He}$ is slightly less than twice that of $^2\text{H}$:

$$4.0026 - 2 \times 2.0141 = -0.0256.$$  

The discrepancy results from the mass equivalent of the energy released in the reaction $2^2\text{H} \rightarrow ^4\text{He}$. From the formula $E = mc^2$, the energy released is:

$$E = mc^2 = 0.0256 \times 1.66 \times 10^{-27} \times (3 \times 10^8)^2 = 3.82 \times 10^{-12} \text{ J} = 23.8 \text{ MeV}.$$  

Thus we obtain an energy that is enormously bigger than the energies of chemical reactions.

### 4.4.6 Energy expressed in terms of momentum

In Newtonian physics, one often writes the kinetic energy as $K = p^2/2m$. In relativity too, it is often convenient to express the energy in terms of the momentum. Let us start from
the formula for the total energy:

\[ E = mc^2 = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}. \]

Now square both sides:

\[ E^2 = \frac{m_0^2c^4}{1 - v^2/c^2} = m_0^2c^4 + \left( \frac{m_0^2c^4}{1 - v^2/c^2} - m_0^2c^4 \right) \]

\[ = m_0^2c^4 + \frac{1}{1 - v^2/c^2} = m_0^2c^4 + m^2v^2c^2 \]

\[ = m_0^2c^4 + p^2c^2. \quad (4.42) \]

So the total energy of the body can be expressed in terms of its rest energy \( m_0c^2 \) and its momentum as:

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

### 4.4.7 The energy-momentum four-vector

From eq. 4.42, we have:

\[ E^2 - p^2c^2 = m_0^2c^4. \]

That is, we have another invariant which looks a bit like the magnitude of a vector. In fact it is possible to show if we put the energy and momentum together as follows:

\[ p = \begin{pmatrix} p_x \\ p_y \\ p_z \\ E/c \end{pmatrix} \]

then \( p \) transforms under boosts in the same way as

\[ \begin{pmatrix} x \\ y \\ z \\ ct \end{pmatrix} \]

and that therefore the energy and momentum of a body form a four-vector - the **energy-momentum four-vector** or **four-momentum**. To see this, consider the four-momentum of a particle at rest in frame \( S \):

\[ p = \begin{pmatrix} p_x \\ p_y \\ p_z \\ E/c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ m_0c \end{pmatrix} \]

and the same particle in frame \( S' \), boosted with speed \( v \) along the \( x \) axis:

\[ p' = \begin{pmatrix} p'_x \\ p'_y \\ p'_z \\ E'/c \end{pmatrix} = \begin{pmatrix} -\gamma m_0v \\ 0 \\ 0 \\ \gamma m_0c \end{pmatrix} \]
If the energy-momentum is really a four-vector, they must satisfy

\[
\begin{pmatrix}
p_x' \\
p_y' \\
p_z' \\
E'/c
d\end{pmatrix} = \begin{pmatrix}
\gamma & 0 & 0 & -\gamma v/c \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
-\gamma v/c & 0 & 0 & \gamma
d\end{pmatrix} \begin{pmatrix}
p_x \\
p_y \\
p_z \\
E/c
d\end{pmatrix}
\] (4.43)

and substituting in the values above and doing the matrix multiplication, we see that indeed they do.

Using the rule for scalar products in space-time, the magnitude of the four-vector is

\[
p^2 = p_x^2 + p_y^2 + p_z^2 - E^2/c^2 = p^2 - E^2c^{-2} = -m_0^2c^2
\]

Energy and momentum conservation can be expressed in the fact that **four-momentum is conserved**. Coupled with the fact that the magnitude of four-momentum is invariant, this is the key to relativistic kinematics.

**Aside:** In many treatments the energy-momentum four-vector is defined as \((E, cp_x, cp_y, cp_z)\), that is Energy (and time) are is the zeroth component, not the fourth. And the metric is defined as

\[
G = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
d\end{pmatrix}
\]

Don’t be confused if you see this in books or later courses - it is just a trivial change of convention, done to make the length of the four-vector equal to (positive) \(m^2c^4\).

### 4.4.8 Massless particles

From the expression for total energy \(E\) in terms of momentum \(p\):

\[
E = \sqrt{m_0^2c^4 + p^2c^2},
\]

we see that it is possible for a particle to have energy even if its rest-mass \(m_0\) is zero. If this happens, we have:

\[
E = |p|c.
\]

In order for \(E\) and \(p\) not to be both zero, the massless particle must travel at the speed of light. This is because \(E\) is also given by:

\[
E = mc^2 = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}.
\]

If \(m_0 = 0\), then \(E = 0\), unless \(v = c\). Equivalently, if \(v = c\), we must have \(m_0 = 0\), otherwise we have infinite energy. So a particle can only travel at the speed of light if it is massless, and if it is massless it can only travel at the speed of light.

The most common massless particle is the photon. According to quantum mechanics, photons have an energy \(E\) and momentum \(p\) given by Einstein’s formula:

\[
E = hf = hc/\lambda,
\]

123
and the de Broglie formula:

\[ |\mathbf{p}| = \frac{h}{\lambda} . \]

From these two formulae, we have \( E = pc \), in agreement with what relativity says. The energy momentum four-vector for a massless particle is still

\[
\left( \begin{array}{c}
px \\
py \\
Hz \\
E/c
\end{array} \right)
\]

but since \( \mathbf{p}^2 = E^2/c^2 \), the magnitude of the four-vector is zero.