Special Relativity

C.R. Prior
Trinity College, University of Oxford, and Rutherford Appleton Laboratory, U.K.

1 Introduction

At the turn of the twentieth century, the development of the Special Theory of Relativity brought into question many of the ideas in classical mechanics that had previously been regarded as fundamental. Maxwell's equations governing Electromagnetism had been formulated 30-40 years earlier and, although it was not appreciated until later, were to turn out to be closely interlinked with the new theory and supplied convincing evidence for its eventual acceptance. In fact it was Lorentz who laid the groundwork for relativity through his studies of electrodynamics, while Einstein contributed crucial concepts and placed the theory on a consistent and general footing. Beyond this, work throughout the twentieth century demonstrated that, even though its origins might have lain in electromagnetism and optics, Special Relativity can be applied to all types of interaction except large-scale gravitational phenomena. In modern physics, the theory serves as a benchmark for descriptions of the interactions between elementary particles, and relativistic features are now so well established that they form basic criteria to be built into any new theory.

This paper starts with a brief description of the experimental basis for Special Relativity, followed by a more detailed derivation of the mathematical ideas behind its structure. The Lorentz transformation and its consequences are covered, with worked examples. The concept of space–time is discussed and leads to the 4-vector formalism which underlies the theory. Modifications to fit classical mechanics into the new framework are described. Topics related to accelerators, such as the connections between the energy, momentum and velocity of particles, are presented, and a final example, looking at particle interactions from different frames, is included for its relevance to colliding beams.

2 Historical background and key experiments

Historically, the turn of the twentieth century was a crucial time in reconciling inconsistencies between ideas in electromagnetism and optics and the fundamental laws of mechanics. A wave theory based on Maxwell's equations had previously been shown to correlate electromagnetism and optics, but assumed the existence of a medium (the ether)—of negligible density, permeating all space with negligible interaction with matter—in which light could propagate. It was also known that the laws of mechanics were the same in different coordinate systems moving uniformly relative to each other, i.e., invariant under Galilean coordinate transformations. But if the ether existed, the laws of electromagnetism could not be invariant under Galilean transformations, so they could only hold in a preferred coordinate system where the ether was at rest. In this system the velocity of light in vacuum was equal to a quantity labelled c (2.99792458×10^8 m/s and by implication that it could not be equal to c in other coordinate frames.

Several attempts were made to reconcile electromagnetism with the rest of physics. Various suggestions were put forward, for example: that the velocity of light is equal to c in coordinate systems in which the source is at rest; that the preferred reference frame for light is the coordinate system in which the medium through which the light is propagating is at rest; or that the ether has a very small interaction with matter, sufficient to be carried along with astronomical bodies such as the earth.

Experiments brought the demise of these ideas and ultimately led to the birth of Special Relativity. The three most fundamental are:

(i) *The aberration of star light.* The small shift in the apparent position of distant stars during the year was recorded in ancient times and can be simply explained by the motion of the earth in its orbit around the sun (at a velocity $\sim 3 \times 10^4$ m/s). This explanation contradicts the hypothesis that the velocity of light is determined by the transmitting medium (our atmosphere) or that the ether is dragged along by the earth. In neither case would aberrations occur.

(ii) *Fizeau's experiments* measured the velocity of light in a swiftly moving liquid in a pipe, first in the direction of and then opposed to the propagation of light. His results were not consistent with any previous assumptions, and could only be made so if it was assumed that bodies smaller than the earth could carry the ether with them in an artificial way involving their refractive index.

(iii) *The Michelson–Morley experiment* was specifically aimed at detecting a motion of the earth relative to the ether at rest, where the velocity of light is c. Light rays were transmitted along paths both parallel and perpendicular to the direction of motion of the earth and reflected back to the observer from silvered mirrors. The expected small differences in the times taken to traverse the paths were not detected and, although the experiment has subsequently been repeated many times with various modifications, no evidence for relative motion through the ether has ever been found.

Although the negative result of the Michelson–Morley experiment can be explained by the ether-drag hypothesis, that hypothesis is inconsistent with the aberration of starlight. Only theories where the velocity of light is constant relative to the source (known as 'emission theories') are in accord with (i), (ii) and (iii), but other experiments exclude these proposals as well. Various alternatives were conceived, notable amongst which was the suggestion by FitzGerald and Lorentz that the null result obtained by Michelson–Morley could be explained while retaining the ether concept if all material objects are contracted in their direction of motion as they move through the ether. The rule of contraction is

$$L(v) = L_0 \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}} .$$

The ether advocates were really clutching at straws but the idea cannot be dismissed and does in fact contain the germs of Special Relativity.

3 The postulates of Special Relativity

Two basic ideas are important in the structured formulation of Special Relativity, helping to explain where Newton went wrong and how new thinkers, such as Einstein, Minkowski and Lorentz put the theories to rights. First, we have the idea of simultaneity, implicit in the statement that two clocks at points A and B are said to be *synchronized* if they read the same time at the mid-point of AB. Secondly, there is the concept of an *inertial frame*, defined to be a frame in which particles acting under no forces move with constant velocity.

Using ideas from projective geometry, it is fairly easy to prove that transformations between such frames must be linear. More formally: *The time and position coordinates* (t, x, y, z) *of a particle with respect to a frame of reference* F *are linearly related to those* (t', x', y', z') *in another frame* F'*, the frames both being inertial.*

Thus, if we consider only transformations in t and x, there must be constants $\alpha, \beta, \gamma, \delta$ such that $t' = \alpha t + \beta x$, $x' = \gamma t + \delta x$. Consider a point fixed in F (i.e., x fixed as t varies). Then $\mathrm{d}x' = \gamma\,\mathrm{d}t,\ \mathrm{d}t' = \alpha\,\mathrm{d}t$ and so

$$\frac{\gamma}{\alpha} = \frac{\mathrm{d}x'}{\mathrm{d}t'} = \text{velocity of frame } F \text{ with respect to } F' = v(F, F') . \tag{1}$$

If instead one takes a point fixed in F' (i.e., x' fixed as t' varies), one has $\mathrm{d}t' = \alpha\,\mathrm{d}t + \beta\,\mathrm{d}x,\ 0 = \gamma\,\mathrm{d}t + \delta\,\mathrm{d}x$. Thus

$$-\frac{\gamma}{\delta} = \frac{\mathrm{d}x}{\mathrm{d}t} = \text{velocity of frame } F' \text{ with respect to frame } F = v(F', F) . \tag{2}$$

One would expect $v(F, F') = -v(F', F)$ so that $\alpha = \delta$ (see below).

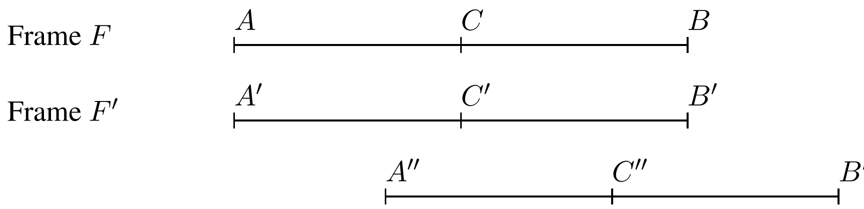
Practically, we can only consider relations between inertial frames such that our measuring apparatus (e.g., rulers and clocks) can actually be transferred from one to another. Such frames are said to be *related*. To go further we need two additional assumptions, that:

(1) the behaviour of apparatus transferred from F to F' is independent of the mode of acceleration.

(2) apparatus transferred from F to F' and then from F' to F'' agrees with apparatus transferred directly from F to F''.

With these assumptions and definitions, it is possible to state **The Principle of Special Relativity**: that all physical laws take equivalent forms in related inertial frames, so that we cannot distinguish between the frames.

Even in the 1900s, this was hardly new. Newton was aware of it, but he based his mechanics on the two fundamental premises (a) a rigid body has the same size in all frames, and (b) time is absolute. However, a very simple thought experiment shows why a revision of these ideas was needed. Consider two points A and B in an inertial frame F. Two events can be said to be simultaneous in F if light rays emitted from A and B at the time of the event meet at the mid-point C of AB.



Suppose a second frame F' moves with velocity v relative to frame F. The diagram shows that by the time the light rays meet at C, C' will have moved to $C'' \neq C$, so that events which are simultaneous in F cannot be simultaneous in F'. We conclude that simultaneity is not absolute but depends on the frame of reference under consideration.

Einstein's reformulation adopted new postulates more in line with these observations. Instead of Newton's hypotheses, he assumed (a) the velocity of light is finite, and (b) the velocity of light has the same value in any inertial frame. These two assumptions lie at the basis of the theory of Special Relativity.

4 The special Lorentz transformation

The negative results of the Michelson–Morley and related experiments led to the formulation of a new theory based on Einstein's two postulates (a) and (b). Let F and F' be two inertial frames of reference equipped with synchronised clocks such that, when $t = t' = 0$, the spatial origins coincide at O. A flash of light, emitted from O at $t = 0$ becomes, in frame F at time t, $ct = \sqrt{x^2 + y^2 + z^2}$, and in frame F' at time t' becomes $ct' = \sqrt{x'^2 + y'^2 + z'^2}$ since c is the same in both F and F'. We demand that these coincide. Thus

$$\left.\begin{array}{rrcl} & P \equiv x^2 + y^2 + z^2 - c^2t^2 & = & 0 \\ \text{whenever} & Q \equiv x'^2 + y'^2 + z'^2 - c^2t'^2 & = & 0 \end{array}\right\} . \tag{3}$$

According to the theorem above, x', y', z', t' are linear functions of x, y, z, t, so that Q is quadratic in x, y, z, t. We therefore have two quadratic functions, P and Q, of the same variables which vanish at the same points. This is only possible if

$$P = kQ \tag{4}$$

where k is independent of x, y, z, t.

Within each frame of reference we can rotate the coordinate axes until Ox and $O'x'$ are both parallel to the direction of relative motion, Oy is parallel to $O'y'$, and Oz is parallel to $O'z'$. This leaves invariant the forms of P and Q. Since the motion can at most produce a re-scaling of lengths in the two-directions Oy and Oz, the transformation must be of the form

$$\begin{aligned} t' &= \alpha t + \beta x \\ x' &= \gamma t + \delta x \\ y' &= \epsilon y \\ z' &= \zeta z \,. \end{aligned} \tag{5}$$

From Eq. (4), we deduce

$$\frac{1}{k}\left(x^2 + y^2 + z^2 - c^2 t^2\right) = \frac{1}{k} P = Q = (\gamma t + \delta x)^2 + \epsilon^2 y^2 + \zeta^2 z^2 - c^2(\alpha t + \beta x)^2 \,,$$

and equating coefficients of individual variable terms gives

$$\epsilon^2 = \zeta^2 = \frac{1}{k} \tag{6}$$

$$\delta^2 - c^2\beta^2 = \frac{1}{k} \tag{7}$$

$$\gamma^2 - c^2\alpha^2 = -\frac{c^2}{k} \tag{8}$$

$$c^2\alpha\beta - \gamma\delta = 0 \,. \tag{9}$$

From Eqs. (7) and (8),

$$c^2\alpha^2\left(c^2\beta^2 + \frac{1}{k}\right) = c^2\alpha^2\delta^2 = \delta^2\left(\gamma^2 + \frac{c^2}{k}\right) \,,$$

and so, invoking Eq. (9),

$$\alpha^2 = \delta^2 \,.$$

However, we saw in Eqs. (1) and (2) that $v(F, F') = \gamma/\alpha$ and $v(F', F) = -\gamma/\delta$ and argued that it is natural to make the choice $\alpha = +\delta$. With this substituted in Eq. (9), we find that

$$\beta = \frac{\gamma}{c^2} = -\frac{\alpha v}{c^2} \qquad \text{where } v = v(F', F) \,. \tag{10}$$

Now, from Eq. (8),

$$\frac{1}{k} = \alpha^2\left(1 - \frac{v^2}{c^2}\right) \qquad \Longleftrightarrow \qquad \alpha = \pm\frac{1}{\sqrt{k}\sqrt{(1 - v^2/c^2)}} \,. \tag{11}$$

Choosing the positive root preserves the sense of time and one similarly takes positive roots of Eq. (6). Thus the transformation Eq. (5) is of the form

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

Although k is independent of x, y, z, t, it seems reasonable to suppose that it depends on $\mathbf{v}$. However, because of the isotropy of space, it cannot depend on the direction of the relative motion, only on its magnitude. Further, a transformation from F to F' followed by the reverse transformation from F' back to F would be expected to lead to the identity. Thus

$$1 = k(\mathbf{v})k(-\mathbf{v}) = k(|\mathbf{v}|)k(|\mathbf{v}|) = (k(v))^2 \ .$$

We deduce that $k(v) = 1$ and finally arrive at the **Special Lorentz Transformation**:

$$\left.\begin{array}{rcl} t' & = & \gamma\left(t - \dfrac{vx}{c^2}\right) \\ x' & = & \gamma(x - vt) \\ y' & = & y \\ z' & = & z \end{array}\right\} \qquad \text{where} \qquad \gamma = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \ . \tag{12}$$

There is the following, more general, form, which we note, valid when the relative motion of the frames is not parallel to a coordinate axis:

$$\begin{array}{rcl} \mathbf{x}' & = & \mathbf{x} + \mathbf{v}\left(\gamma t + (\gamma - 1)\dfrac{\mathbf{v}\cdot\mathbf{x}}{v^2}\right) \\ t' & = & \gamma\left(t + \dfrac{\mathbf{v}\cdot\mathbf{x}}{c^2}\right) \ . \end{array} \tag{13}$$

4.1 Consequences of the Lorentz transformation

Consider first a rigid rod in F' and lying along the x'-axis between points A and B. Its length as measured in F' is

$$L' = x'_A - x'_B \tag{14}$$

independent of the time t' at which we measure it. With respect to F the rod is moving and it only makes sense to talk about its length if we measure the position of its ends at exactly the same time. At the instant t in F at which these ends occupy positions x_A and x_B, we have, by Eq. (12),

$$x'_A = \gamma(x_A - vt) \ , \qquad x'_B = \gamma(x_B - vt)$$

so that

$$L' = \gamma(x_A - x_B) = \gamma L > L \ . \tag{15}$$

The length of the bar accordingly suffers contraction when it is moved longitudinally relative to an inertial frame. This is the Fitzgerald contraction, and is not to be thought of as the physical reaction of the rod to its motion (cf. the contraction of a metal rod when cooled) but rather as due to the changed relationship between the rod and the instruments measuring its length: some instruments are stationary with respect to the bar, others are moving with respect to it. Also the measurement of L' can be carried out without the assistance of a clock, but the second operation involves simultaneous observation of the two ends of the bar and clocks must be employed. It is the procedure in the measurement that actually *defines* the length.

Now consider two events occuring at the same point (x, y, z) of frame F and different times t_A and t_B, as measured in F. Observers with synchronised clocks in F' will measure the time interval as

$$\Delta t' = t'_B - t'_A = \gamma(t_B - t_A) = \gamma\Delta t \tag{16}$$

using Eq. (12). This equation shows that relative to F' the clock moving with F will appear to have its rate reduced by a factor $1/\gamma$. This is the *time dilatation* effect. It implies that all physical processes will evolve more slowly when observed from a frame relative to which they are moving. Thus the rate of decay of cosmic rays moving with high velocities relative to the earth has been observed to be reduced by exactly the factor predicted by Eq. (16). In particle accelerators, rapid acceleration to high velocities can be used to extend the laboratory lifetime of muon beams, for example, and this technique lies behind current ideas for a muon-based neutrino factory or a muon collider.

It may also be deduced that if a human passenger were launched at high speed from the earth and after proceeding a great distance were to return at the same high speed, observations made from the earth would indicate that all physical processes within the rocket, including the ageing of the passenger, would be retarded. As all processes would be equally affected, the passenger would be unaware of this effect, but nevertheless, upon his return to earth he would find that his estimate of the duration of flight was less than the terrestrial estimate. One might also claim that the passenger is entitled to regard himself as at rest and the earth as having suffered the displacement, so that the terrestrial estimate should be less than his own. This, *the clock paradox,* is resolved by observing that a frame moving with the rocket is subject to an acceleration relative to an inertial frame and consequently cannot be regarded as inertial. Since the results of Special Relativity apply only to inertial frames, the rocket passenger is not justified in making use of them in his own frame.

4.2 Examples

Example 1: *A rocket passes at speed v through a tunnel of length L. Observer B is in the tail of the rocket and observer A is stationed in the nose. Their clocks are synchronized and they are a distance L apart in the rocket. Two other observers, X and Y, are positioned at the tunnel exit and entrance respectively, also with synchronized clocks. The following events occur:*

1. *X sees the rocket nose (and A) emerge from the tunnel.*
2. *Y sees the rocket tail (and B) disappear into the tunnel.*

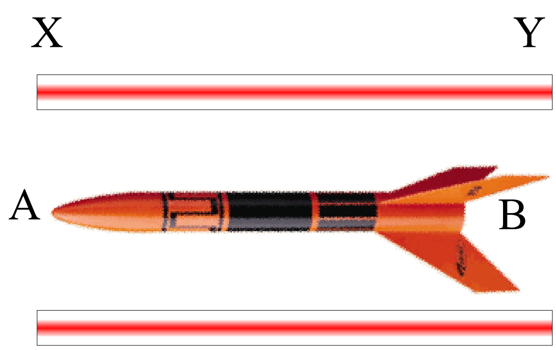
(i) If X's clock read zero at event (1), what did Y's clock read at event (2)?

(ii) If A's clock read zero at event (1), what did B's clock indicate at event (2)?

(iii) Where was B when his clock indicated zero?

(iv) Where was A when his clock indicated the same as B's at event (2)?

The essence of this problem is that X and Y see the moving rocket as Lorentz contracted to L/γ and therefore shorter than the tunnel. On the other hand, to A and B the tunnel is moving and it is contracted to L/γ so their rocket is longer.



(i) Since the clocks are synchronized, if X's clock reads zero at event (1), then so does Y's and at this time, Y will claim that the tail of the (contracted) rocket is already inside the tunnel by a distance $L - L/\gamma$. He will therefore say that, when the tail passed him, his clock read

$$-\frac{L}{v}\left(1-\frac{1}{\gamma}\right). \tag{Ex1.i}$$

(ii) Similarly, if the exit of the tunnel (X) coincides with A at time zero, since the observers in the rocket see the tunnel contracted, B will claim he is still a distance $L - L/\gamma$ outside the entrance and that when he gets there his clock will read

$$+\frac{L}{v}\left(1-\frac{1}{\gamma}\right). \tag{Ex1.ii}$$

(iii) When B's clock read zero, A's clock also read zero and the front of the rocket was just emerging from the tunnel. B will say he still has a distance $L - L/\gamma$ to travel before he enters. This is in his frame; in the frame of the tunnel, the distance becomes

$$\gamma \times \left(L - \frac{L}{\gamma}\right) = L(\gamma - 1)\ . \tag{Ex1.iii}$$

(iv) Similarly, at event (2), B is just entering the tunnel, and because it is contracted, A is a distance $L - L/\gamma$ outside. Converted to the tunnel frame, this means that the front of the rocket has left the tunnel and is a distance $L(\gamma - 1)$ down the track.

Though puzzling, the results are quite consistent when one understands how the length of a moving object is defined. In this example, the heuristic approach using length contraction is acceptable, but in more complicated scenarios it may be necessary to work within the safety of the mathematical Lorentz formulation, Eq. (12). In this case $F(t, x)$ would be the frame of X and Y and $F'(t', x')$ would be the frame of A and B. Event (1) is $(x, t) = (0, 0)$, at event (2) $x = L$, at A $x' = 0$ and at B $x' = L$. The transformation formulae are

$$x = \gamma(x' - vt'), \qquad t = \gamma\left(t' - \frac{vx'}{c^2}\right) \tag{Ex1.iv}$$

$$x' = \gamma(x + vt), \qquad t' = \gamma\left(t + \frac{vx}{c^2}\right) \tag{Ex1.v}$$

since the rocket in the picture moves from right to left. Thus to answer part (i), we put $x = x' = L$ into Eq. (Ex1.v) to deduce t as in Eq. (Ex1.i) above. For part (ii), we put these values into Eq. (Ex1.iv) to deduce t' as in Eq. (Ex1.ii). For part (iii), $t' = 0$, $x' = L$ gives $x = \gamma L$, or $(\gamma - 1)L$ outside the tunnel, as in Eq. (Ex1.iii); and for part (iv) we put $x' = 0$, $t' = \dfrac{L}{v}\left(1 - \dfrac{1}{\gamma}\right)$ to get $x = -L(\gamma - 1)$. ◊◊

The following example, concerning the change in frequency measured by a moving observer, is the relativistic counterpart of the Doppler shift.

Example 2: *Using the Lorentz transformation, find an expression for the frequency ν' observed by an observer O' when light of frequency ν is emitted from a point O moving directly away from O' with velocity v.*

Let F and F' be inertial frames with parallel coordinate axes centred on O and O', respectively, such that the relative motion is directed along Ox. Successive light pulses emitted from O in F are represented by the two events $(t_1, 0, 0, 0)$ and $(t_2, 0, 0, 0)$ where $t_2 - t_1 = 1/\nu$. By Eq. (12) and the observation that F' has velocity $-v$ along Ox relative to F, the events correspond in F' to $\gamma(t_1, vt_1, 0, 0)$

and $\gamma(t_2, vt_2, 0, 0)$. But light signals in F' emitted at x' reach O' a time x'/c later. Thus the pulses are received by the observer O' at times $\gamma t_i + \gamma v t_i/c$ for $i = 1, 2$. He deduces a frequency given by

$$\frac{1}{\nu'} = \gamma\left(1 + \frac{v}{c}\right)(t_2 - t_1) = \gamma\left(1 + \frac{v}{c}\right)\frac{1}{\nu} \, .$$

Thus, from Eq. (12),

$$\nu' = \nu\left[\frac{c - v}{c + v}\right]^{\frac{1}{2}} \, . \tag{Ex2.i}$$

◊◊

5 Space–time

In Section 4 it was proved that, since $k = 1$, the quantity P given by Eq. (3) is invariant, i.e., has the same value for all observers employing inertial frames and rectangular coordinate axes. With respect to a general origin of coordinates (x_0, y_0, z_0) and origin of time t_0, this quantity is

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \tag{17}$$

where $\Delta x = x - x_0$ etc. The 4-dimensional space with coordinates (t, x, y, z) is called *space–time* and the point (t, x, y, z) or $(t, \mathbf{x})$ is called *an event*. Δs is referred to as the separation between the two events $(t, \mathbf{x})$ and $(t_0, \mathbf{x_0})$. The path of a succession of events in space–time is called *the world-line*.

The *proper time* τ between two events is defined by

$$\Delta\tau^2 = \frac{1}{c^2}\,\Delta s^2 \, . \tag{18}$$

Calling Δd the distance $|\mathbf{x} - \mathbf{x_0}|$, we have

$$\Delta\tau^2 = \Delta t^2 - \frac{1}{c^2}\,\Delta d^2 \, . \tag{19}$$

Suppose now that a new inertial frame F' is defined, moving in the direction of the line joining the two events with speed $\Delta d/\Delta t < c$. Relative to F', the events occur at the same point and hence $\Delta d' = 0$. By Eq. (19) therefore

$$\Delta\tau = \Delta t'$$

and one deduces that the proper time interval between two events is the ordinary time interval measured in a frame in which the events occur at the same point (if it exists). Then $\Delta\tau^2 > 0$ and the separation is termed *timelike.*

If, on the other hand, it is possible to find a frame F' relative to which the events are simultaneous, $\Delta t' = 0$ and

$$\Delta\tau^2 = -\frac{1}{c^2}\Delta d'^2 < 0 \, ,$$

and $\Delta d/\Delta t > c$. The separation is now called *spacelike*.

If the separation is timelike, $\Delta d/\Delta t < c$ and it is possible for a material body to be present at both events, but this is not true for a spacelike separation where $\Delta d/\Delta t > c$. The intermediate case, when $\Delta d/\Delta t = c$ and $\Delta\tau = 0$ corresponds to a *null* or *lightlike* separation and only a light pulse can be present at both events. It may also be observed that the proper time interval between the transmission and the receipt of a light signal is zero.

6 Four-vectors, invariants and covariance

A physical quantity which has the same numerical value for all observers is called an invariant or 4-scalar. Examples are the separation of two events, the phase of a wave, and the rate of radiation of a moving charged particle.

From the discussion so far it is already apparent that in Special Relativity the concepts of space and time are intertwined. To treat the subject rigorously would require definitions of tensors, metric and covariant and contravariant vectors. Fortunately for accelerator physicists' purposes, it is sufficient to adopt a simpler approach.

Define the *position 4-vector* to be the set of four quantities given by

$$\mathcal{X} = (ct, \mathbf{x}) \ . \tag{20}$$

$\mathcal{X}$ consists of two parts, time and the normal position 3-vector. Under a Lorentz transformation, its components change according to Eq. (12), which we can write in matrix form as

$$\begin{bmatrix} ct' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \gamma & -\dfrac{\gamma v}{c} & 0 & 0 \\ -\dfrac{\gamma v}{c} & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix} . \tag{21}$$

We denote the matrix by Λ so that the transformation law can be written compactly as $\mathcal{X}' = \Lambda \mathcal{X}$.

Any physical quantity, such as $\mathcal{X}$, with four components which transform under Λ as in Eq. (21), is called a *4-vector*, and equations involving 4-vectors hold in all inertial frames. For example, if $\mathcal{A}$ and $\mathcal{B}$ are 4-vectors and $\mathcal{A} = \mathcal{B}$ in one frame, then $\Lambda\mathcal{A} = \Lambda\mathcal{B}$, so $\mathcal{A}' = \mathcal{B}'$ also holds in the new frame. In classical mechanics the scalar products of 3-vectors are invariant and we would like an analogous result in relativity. Now we know from Eq. (17) that $(ct)^2 - \mathbf{x}\cdot\mathbf{x}$ is invariant; therefore, referring to Eq. (20), we define the relativistic scalar product of $\mathcal{A} = (a_0, \mathbf{a})$ and $\mathcal{B} = (b_0, \mathbf{b})$ by

$$\mathcal{A}\cdot\mathcal{B} = a_0 b_0 - \mathbf{a}\cdot\mathbf{b} \ . \tag{22}$$

Then

$$\mathcal{A}'\cdot\mathcal{B}' = \mathcal{A}\Lambda^{\mathrm{T}}\Lambda\mathcal{B} = \mathcal{A}\cdot\mathcal{B} \qquad \text{since} \qquad \Lambda^{\mathrm{T}}\Lambda = I \ . \tag{23}$$

With this definition therefore, the scalar product of two 4-vectors is invariant.

7 Special Relativity mechanics

In Sections 4.1 and 5 it was shown that the time interval between two events is dependent on the frame of reference from which the events are observed. The proper time interval $\mathrm{d}\tau$ is the time interval which would be measured by clocks in the frame for which the events occur at the same point. It is related to the time interval $\mathrm{d}t$ in any other frame by Eq. (16).

$$\mathrm{d}t = \gamma\,\mathrm{d}\tau \ . \tag{24}$$

If a clock leaves a point A at time t_1 and arrives at a point B at time t_2 the time of transit as registered by the moving clock will be

$$\tau_2 - \tau_1 = \int_{t_1}^{t_2} \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}} \mathrm{d}t \ . \tag{25}$$

The successive positions of the clock together with the times it occupies these positions constitute a series of events which lie on the clock's world-line in space–time. If F is an inertial frame of reference and $(t, \mathbf{x})$ and $(t + \mathrm{d}t, \mathbf{x} + \mathrm{d}\mathbf{x})$ represent adjacent points on the world-line in F, the velocity vector of the moving clock with respect to F is

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\,.$$

This, however, does not possess the transformation properties required for a 4-vector in Special Relativity. But a 4-vector with the correct properties can be defined as follows: $\mathrm{d}\mathbf{x}$ is a displacement vector relative to rectangular axes and $\mathrm{d}\tau$ is an invariant. Thus, $\mathrm{d}\mathcal{X}/\,\mathrm{d}\tau$ is a 4-vector relative to Lorentz transformations in space–time. Under a Lorentz transformation, the numerator takes on a factor Λ and the denominator is unchanged. This quantity will be denoted by $\mathcal{V}$ and is called the velocity 4-vector. From Eqs. (24) and (20), it follows that

$$\mathcal{V} = \frac{\mathrm{d}\mathcal{X}}{\mathrm{d}\tau} = \gamma\frac{\mathrm{d}}{\mathrm{d}t}(ct, \mathbf{x}) = \gamma(c, \mathbf{v})\,. \tag{26}$$

Knowing how this transforms enables us to calculate how the components of $\mathbf{v}$ appear when measured from a new frame F': by comparison with Eq. (20) we merely write out the transformation equations Eq. (21) with t replaced by γ and $\mathbf{x}$ replaced by $\gamma\mathbf{v}$. Thus

$$\begin{array}{rcl} \gamma_{v'}v'_x & = & \gamma(\gamma_v v_x - u\gamma_v) \\ \gamma_{v'}v'_y & = & \gamma_v v_y \\ \gamma_{v'}v'_z & = & \gamma_v v_z \\ \gamma_{v'} & = & \gamma(\gamma_v - uv\gamma_v/c^2) \end{array} \tag{27}$$

where $\gamma_v = (1 - v^2/c^2)^{-\frac{1}{2}}$ and $\gamma = (1 - u^2/c^2)^{-\frac{1}{2}}$, u being the relative velocity of the frames F and F'. Eliminating $\gamma_{v'}$, we have the velocity transformation laws:

$$\left.\begin{array}{rcl} v'_x & = & Q(v_x - u) \\ v'_y & = & Qv_y/\gamma \\ v'_z & = & Qv_z/\gamma \end{array}\right\} \qquad \text{with} \qquad Q = \left(1 - \frac{uv_x}{c^2}\right)^{-1}\,. \tag{28}$$

Note that if $v_x = c$, then also $v'_x = c$, confirming that light propagates with speed c in all inertial frames.

Consider now conservation of momentum for the collision of two particles. To generalize the familiar mathematical expression, we tentatively write

$$\sum M\mathcal{V} = \text{constant}$$

where $\mathcal{V}$ is the 4-velocity vector and M (to preserve an overall 4-vector form) represents an invariant associated with the particle in question that is to correspond to its classical mass. By Eq. (26), this implies that

$$\sum m(c, \mathbf{v}) \quad \text{is conserved}$$

where $m = M\gamma$. If m is identified with the relativistic analogue of Newtonian mass, it appears that our tentative conservation law incorporates both the principles of conservation of (3-) momentum and conservation of mass:

$$\sum m\mathbf{v} \text{ is conserved;} \qquad\qquad \sum m \text{ is conserved.}$$

M is called the *rest mass* of the particle and is usually denoted by m_0. Then the relativistic mass is

$$m = m_0\left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} = m_0\gamma\,. \tag{29}$$

m_0 is the mass of a particle in its rest-frame (where $\gamma = 1$) and one must regard the mass of a moving particle as being dependent on its speed. As $v \to c$ inertia effects become increasingly serious and prevent the speed of light being attained by any particle. This is in agreement with observations.

The 4-momentum vector is accordingly defined by

$$\mathcal{P} = m_0 \mathcal{V} \; . \tag{30}$$

Being the product of an invariant and a 4-vector, $\mathcal{P}$ has the desired transformation properties for a vector. Its components are

$$\mathcal{P} = m_0 \gamma (c, \mathbf{v}) = (mc, m\mathbf{v}) = (mc, \mathbf{p}) \tag{31}$$

where $\mathbf{p}$ is the classical momentum.

Newton's second law, $\mathbf{f} = \dfrac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}$, can now be generalized within the framework of Special Relativity. In the classsical form, $\mathbf{f}$ is the force acting on a particle having mass m and velocity $\mathbf{v}$ relative to some inertial frame. It implies that, if equal and opposite forces act upon two colliding particles, momentum is conserved. The conclusion is certainly true, but it turns out that if the forces are equal and opposite for one observer, they are not so in general for another. Accordingly we define the 4-force $\mathcal{F}$ by the equation

$$\mathcal{F} = \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\tau} = m_0 \frac{\mathrm{d}\mathcal{V}}{\mathrm{d}\tau} \; . \tag{32}$$

This has the correct transformation properties for a vector and has components

$$\begin{aligned} \mathcal{F} &= m_0 \gamma \frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} \\ &= \gamma \left(c \frac{\mathrm{d}m}{\mathrm{d}t}, \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \right) \\ &= \gamma \left(c \frac{\mathrm{d}m}{\mathrm{d}t}, \mathbf{f} \right) \; . \end{aligned} \tag{33}$$

From Eq. (26) we calculate

$$\mathcal{V} \cdot \mathcal{V} = \gamma^2 (c^2 - v^2) = c^2 \; . \tag{34}$$

Differentiate with respect to τ:

$$0 = \mathcal{V} \cdot \frac{\mathrm{d}\mathcal{V}}{\mathrm{d}\tau} = \frac{1}{m_0} \mathcal{V} \cdot \mathcal{F} \; . \tag{35}$$

This result, which has important consequences, can be written in component form as

$$c^2 \frac{\mathrm{d}m}{\mathrm{d}t} - \mathbf{v}.\mathbf{f} = 0 \; . \tag{36}$$

By definition, $\mathbf{v}.\mathbf{f}$ is the rate at which the force is doing work, so that during a time interval $[t_1, t_2]$ the work done is

$$\int_{t_1}^{t_2} c^2 \frac{\mathrm{d}m}{\mathrm{d}t} \, \mathrm{d}t = m_2 c^2 - m_1 c^2 \; . \tag{37}$$

Classically, we equate the work done by a force to the change in kinetic energy of the moving particle, T. Hence one must define T by a formula of the form

$$T = mc^2 + \text{ constant } . \tag{38}$$

When $v = 0$, $T = 0$ and so the constant is $-m_0 c^2$. Thus

$$T = mc^2 - m_0 c^2 = m_0 c^2 (\gamma - 1) \; .$$

If v is small, using a binomial expansion,

$$\gamma - 1 = \left(1 - \frac{v^2}{c^2} \right)^{-\frac{1}{2}} - 1 \approx \frac{1}{2} \frac{v^2}{c^2} + \mathrm{O} \left(\frac{v^4}{c^4} \right) \; ,$$

so T approximates to $\frac{1}{2}m_0 v^2$ in agreement with classical theory.

Suppose two equal elastic particles approach each other along the same straight line with equal speeds v. If their rest masses are both m_0, the net mass before collision is $2m_0\gamma$. We accept as a fundamental principle that this mass will be conserved during the collision. However, it is clear that at some instant during the collision both particles will be brought to rest and their masses at this instant will be their rest masses m_0'. By our principle

$$2m_0' = 2m_0\gamma \ ,$$

so that at this instant the rest mass of each particle has increased by

$$m_0\gamma - m_0 = \frac{1}{c^2}T \tag{39}$$

where T is the original kinetic energy of the particle. In losing this kinetic energy the particle has had an equal amount of work done upon it by the force of interaction and this has resulted in a distortion in the elastic material of which it is made. This distortion is a maximum when the particle is at rest and the elastic potential energy as measured by the work done will be exactly T. If we assume that this increase in internal energy of the particle leads to a proportional increase in rest mass, the increment [Eq. (39)] is explained. Considerations such as this suggest strongly that mass and energy are equivalent. All forms of energy, mechanical, thermal, electromagnetic, etc., are to be thought of as possessing inertia of mass m, according to Einstein's equation

$$\boxed{E = mc^2} \ . \tag{40}$$

Written as

$$E = T + m_0c^2 \ , \tag{41}$$

m_0c^2 can be interpreted as the internal energy of the particle when stationary. Such energy would be released if the particle could be completely converted into electromagnetic energy and is the source of energy in an atomic explosion.

8 Relationships between energy, momentum, and velocity

Relativistic kinematics is the standard tool of high-energy physics and we now give some illustrations of the methods used to tackle problems.

Several identities are useful in switching between velocity v, momentum p (which is proportional to γv), and energy (which is effectively γ). In accelerator theory, it is common to write $\beta = v/c$, $0 \le \beta \le 1$. Thus it follows from the definition

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} = (1-\beta^2)^{-\frac{1}{2}} \ , \tag{42}$$

that

$$(\beta\gamma)^2 = \frac{\gamma^2 v^2}{c^2} = \gamma^2 - 1 \ . \tag{43}$$

We also have

$$\beta^2 = \frac{v^2}{c^2} = 1 - \frac{1}{\gamma^2} \ . \tag{44}$$

Charged particles in accelerators usually have energies and momenta spread over a small range of values. By differentiating Eq. (42) to Eq. (44), we see that first-order variations $\Delta\beta$, $\Delta\gamma$ from the means β and γ are related by

$$\beta\Delta(\beta\gamma) = \Delta\gamma = \beta\gamma^3\Delta\beta \ . \tag{45}$$

Since the energy of a particle is $E = m_0c^2\gamma$ and the momentum (for one-dimensional motion) is $p = m_0\gamma v = m_0c\beta\gamma$, we have

$$\frac{\Delta p}{p} = \frac{\Delta(\beta\gamma)}{\beta\gamma} = \gamma^2\frac{\Delta\beta}{\beta} = \frac{1}{\beta^2}\,\frac{\Delta\gamma}{\gamma} = \frac{\gamma}{\gamma+1}\,\frac{\Delta T}{T}\ , \tag{46}$$

with T the kinetic energy. The complete set of relations between first-order increments in p, E, T, β and γ is given in Table 1.

Table 1: Incremental relationships between energy, velocity, and momentum

<table>
<tr><th></th><th>$\frac{\Delta\beta}{\beta}$</th><th>$\frac{\Delta p}{p}$</th><th>$\frac{\Delta T}{T}$</th><th>$\frac{\Delta E}{E} = \frac{\Delta\gamma}{\gamma}$</th></tr>
<tr><td rowspan="2">$\frac{\Delta\beta}{\beta} =$</td><td rowspan="2">$\frac{\Delta\beta}{\beta}$</td><td>$\frac{1}{\gamma^2}\frac{\Delta p}{p}$</td><td rowspan="2">$\frac{1}{\gamma(\gamma+1)}\frac{\Delta T}{T}$</td><td>$\frac{1}{\beta^2\gamma^2}\frac{\Delta\gamma}{\gamma}$</td></tr>
<tr><td>$\frac{\Delta p}{p} - \frac{\Delta\gamma}{\gamma}$</td><td>$\frac{1}{\gamma^2-1}\frac{\Delta\gamma}{\gamma}$</td></tr>
<tr><td>$\frac{\Delta p}{p} =$</td><td>$\gamma^2\frac{\Delta\beta}{\beta}$</td><td>$\frac{\Delta p}{p}$</td><td>$\frac{\gamma}{\gamma+1}\frac{\Delta T}{T}$</td><td>$\frac{1}{\beta^2}\frac{\Delta\gamma}{\gamma}$</td></tr>
<tr><td>$\frac{\Delta T}{T} =$</td><td>$\gamma(\gamma+1)\frac{\Delta\beta}{\beta}$</td><td>$\left(1+\frac{1}{\gamma}\right)\frac{\Delta p}{p}$</td><td>$\frac{\Delta T}{T}$</td><td>$\frac{\gamma}{\gamma-1}\frac{\Delta\gamma}{\gamma}$</td></tr>
<tr><td>$\frac{\Delta E}{E} =$</td><td>$(\beta\gamma)^2\frac{\Delta\beta}{\beta}$</td><td>$\beta^2\frac{\Delta p}{p}$</td><td rowspan="2">$\left(1-\frac{1}{\gamma}\right)\frac{\Delta T}{T}$</td><td rowspan="2">$\frac{\Delta\gamma}{\gamma}$</td></tr>
<tr><td>$\frac{\Delta\gamma}{\gamma} =$</td><td>$(\gamma^2-1)\frac{\Delta\beta}{\beta}$</td><td>$\frac{\Delta p}{p} - \frac{\Delta\beta}{\beta}$</td></tr>
</table>

More useful than using γ and $\mathbf{v}$ is to concentrate on expressions involving energy E and momentum $\mathbf{p}$. Combining Eq. (40) with the expression (31) for the 4-momentum vector, we have

$$\mathcal{P} = (E/c, \mathbf{p})\ . \tag{47}$$

The quantity $\mathcal{P}\cdot\mathcal{P}$ is an invariant. Its value may be calculated from $\mathcal{P} = m_0\mathcal{V}$ and Eq. (34), giving

$$\mathcal{P}\cdot\mathcal{P} = \frac{E^2}{c^2} - |\mathbf{p}|^2 = m_0^2c^2\ . \tag{48}$$

Since $\mathcal{P}$ transforms in exactly the same way as $\mathcal{X}$, from Eq. (12), we can write down the connection between energy and momentum between inertial frames of reference:

$$\begin{aligned} p'_{\mathrm{x}} &= \gamma\left(p_{\mathrm{x}} - \frac{Ev}{c^2}\right) \\ p'_{\mathrm{y}} &= p_{\mathrm{y}} \\ p'_{\mathrm{z}} &= p_{\mathrm{z}} \\ E' &= \gamma(E - vp_{\mathrm{x}})\ . \end{aligned} \tag{49}$$

It is often helpful when dealing with problems involving a number of particles to work in the centre-of-momentum frame (often loosely called the centre-of-mass frame). Since $\mathcal{P}$ is a 4-vector for an individual particle, so too is

$$\sum_{\text{particles}} \mathcal{P} = \left(\frac{1}{c}\sum E, \sum \mathbf{p}\right)\ . \tag{50}$$

The centre-of-momentum frame (COM) is that in which $\sum \mathbf{p} = \mathbf{0}$.

From Eq. (50), the quantity

$$\frac{1}{c^2}(\sum E)^2 - \left(\sum \mathbf{p}\right)^2 \tag{51}$$

is invariant, equal to (the total energy)$^2/c^2$ in the centre-of-momentum frame. This is an enormously useful invariant. A good rule for solving many problems is to start in the laboratory frame, transform to the centre-of-momentum frame, where you carry out the working of the question, then transform your results back to the laboratory frame. The idea is illustrated in the following examples.

Example 3: *Two particles have equal rest mass m_0. Their total energy in the inertial frame in which one of them is at rest is E_1. In the frame in which their velocities are equal in magnitude but opposite in direction, their total energy is E_2. We show that*

$$E_2^2 = 2m_0c^2E_1 \ .$$

Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be the 4-momenta of the two particles. In the frame in which particle 1 is at rest $\mathcal{P}_1 = (m_0c, \mathbf{0})$ and, since the total energy is E_1, $\mathcal{P}_2$ has the form $\left(\frac{1}{c}(E_1 - m_0c^2), \mathbf{p}\right)$ where $\mathbf{p}$ is the 3-momentum of particle 2.

The second frame is the centre-of-momentum frame since the particles have equal rest masses. Moreover Eq. (48) implies that they have equal energies since the magnitude of their momenta is the same. Thus in this frame

$$\mathcal{P}_1 = \left(\frac{E_2}{2c}, \mathbf{p}'\right), \qquad \mathcal{P}_2 = \left(\frac{E_2}{2c}, -\mathbf{p}'\right) \ .$$

Now consider the product $\mathcal{P}_1 \cdot (\mathcal{P}_1 + \mathcal{P}_2)$. This is invariant and has the same value in both frames. Hence

$$m_0c \times \frac{E_1}{c} - \mathbf{0}.\mathbf{p} = \frac{E_2}{2c} \times \frac{E_2}{c} - \mathbf{p}'.\mathbf{0}$$

or

$$2m_0c^2E_1 = E_2^2 \ . \tag{Ex3.i}$$

Note that, by considering the 4-vector product of $\mathcal{P}_1$ and $\mathcal{P}_1 + \mathcal{P}_2$, we create enough zero terms to eliminate the unknown quantities, $\mathbf{p}$ and $\mathbf{p}'$, which we are not asked to find. ◊◊

The next example makes use of this result.

Example 4: *In an accelerator a proton P_1 with rest mass m collides with an anti-proton P_2 (with the same rest mass), producing two particles W_1 and W_2 with equal mass $M = 100m$. First the experiment takes place with P_1 and P_2 having equal and opposite velocities in the laboratory frame. Calculate the minimum energy E_0 the laboratory had to supply to P_2 in order for W_1 and W_2 to be produced.*

Next the experiment takes place with P_1 at rest. Calculate the minimum energy E_0' the laboratory has to supply to P_2 in order for W_1 and W_2 to be produced in this case, to within 1%.

In the COM frame, since the rest masses are the same and the 3-momenta must be equal and opposite, Eq. (48) shows that the energies of the proton and the antiproton must be equal. Hence the 4-momenta before the collision have the form

$$\mathcal{P}_1 = \left(\frac{E}{c}, \mathbf{p}\right), \qquad \mathcal{P}_2 = \left(\frac{E}{c}, -\mathbf{p}\right) \ .$$

After the collision, when the W-particles are produced, the total 3-momentum must be conserved, so the energies are again the same and

$$\mathcal{P}_{W_1} = \left(\frac{\tilde{E}}{c}, \tilde{\mathbf{p}}\right), \qquad \mathcal{P}_{W_2} = \left(\frac{\tilde{E}}{c}, -\tilde{\mathbf{p}}\right) .$$

Conservation of 4-momentum gives

$$\mathcal{P}_1 + \mathcal{P}_2 = \mathcal{P}_{W_1} + \mathcal{P}_{W_2} .$$

Hence, equating the energy parts, we have

$$E = \tilde{E} \geq \text{rest energy of a } W\text{-particle} = M_0 c^2 = 100 m_0 c^2 . \tag{Ex4.i}$$

In the laboratory frame, the proton is at rest and the antiproton moves with relativistic energy E'. The total energy, $E_1 = E' + m_0 c^2$, is the same for the W-particles produced after the collision. Transformed into the COM frame, this total energy is E_2 given by Eq. (Ex3.i). Thus

$$\begin{aligned} 2m_0c^2(E' + m_0c^2) = 2m_0c^2E_1 &= E_2^2 = (2\tilde{E})^2 \\ &\geq (2M_0c^2)^2 = 4 \times 10^4 (m_0c^2)^2 . \end{aligned}$$

We deduce that

$$E' \geq (2 \times 10^4 - 1) m_0 c^2 \approx 2 \times 10^4 m_0 c^2 ,$$

demonstrating that considerably more energy is required to produce an event via a fixed-target collision ($E_0' = 20000 m_0 c^2$) than with two colliding beams ($E_0 = 100 m_0 c^2$). ◊◊