The Special Theory of Relativity

Vasant Natarajan and Diptiman Sen

Vasant Natarajan is at the Department of Physics, IISc, Bangalore. His current research involves trapping of atoms to carry out high precision tests of fundamental physics. He has earlier worked on high precision mass spectrometry and on the focussing of atomic beams by laser fields.

Diptiman Sen works in the Centre for High Energy Physics, Indian Institute of Science, Bangalore. He is interested in condensed matter systems in low dimensions and quantum field theory.

Keywords
Lorentz transformations, time dilation, $E=mc^2$, fission and fusion, antiparticles.

1. Introduction

The year 2005 has been declared by the United Nations as the International Year of Physics. This has been done to commemorate Albert Einstein's "miraculous year" a century ago, during which the young Einstein made a number of remarkable discoveries that changed the course of physics. In that single year, he completed his PhD thesis and published five papers on three different subjects, namely, the photoelectric effect, Brownian motion, and the special theory of relativity[1].

The two papers that laid out the foundations of the special theory of relativity were published in the journal *Annalen der Physik*, and were titled 'On the electrodynamics of moving bodies' (Vol.17, pp.891–921) and 'Does the inertia of a body depend upon its energy-content?' (Vol.18, pp.639–641). The first paper derived the laws of transformation of spacetime coordinates and electromagnetic fields between two inertial frames[2]. The second paper derived the relation $E = mc^2$, which is perhaps the most famous and widely-recognized equation today.

2. Historical Background

In order to set Einstein's work in proper perspective, recall that, in 1865, Maxwell had unified electricity and magnetism by writing down his equations of electromagnetism. It was soon realized that these equations supported wave-like solutions in a region free of electrical charges or currents (such a region is called the 'vacuum'). Experiments by Hertz and others showed conclusively that these electromagnetic waves (of the right frequency) were to be identified with *light*. Thus, Maxwell's equations brought optics into its unified domain and, furthermore, predicted that light waves should propagate at a finite speed c (about 300,000 km/s) calculable from these equations. With their Newtonian ideas of absolute space and time firmly entrenched, most physicists thought that this speed was correct only *in one special frame*. This special frame was postulated to be in a state of absolute rest, and it was thought that electromagnetic waves were supported by an unseen medium called the *ether*, which is at rest in this frame.

With this picture in mind, it is clear that if the Earth moves with respect to this special frame at a velocity $\vec{v}$, then the speed of light with respect to the Earth can take any value from $c - |\vec{v}|$ to $c + |\vec{v}|$, depending on the particular direction of $\vec{v}$. This is indeed our intuitive understanding of how two velocities add. For example, if two objects have speeds v_1 and v_2 with respect to some fixed frame, then we expect that their relative speed will be $v_1 - v_2$ if they are moving in the same direction, and $v_1 + v_2$ if they are moving in opposite directions. We have developed this 'intuition' based on our everyday experiences, and we have a gut feeling that it is correct. However, we will see below that the correct relativistic law of 'addition of velocities' is somewhat different.

There were thus several attempts to detect the presence of this hypothetical ether or to show that the speed of light was frame dependent. The most famous experiment was done in 1887 by A A Michelson and E W Morley. They used an interferometer to study the speed of light along different directions and at different times of the year (when the orbital velocity of the Earth pointed along different directions). They concluded that *the speed of light with respect to the Earth is always the same and does not depend on the relative motion of the Earth and the light wave*. This observation proved that Maxwell's equations are valid in all (inertial) frames, and

[1] For further details, see the article 'Einstein's miraculous year' by Vasant Natarajan, V Balakrishnan and N Mukunda, *Resonance*, Vol.10, No.3, pp.35–56, 2005.

[2] An inertial frame is a set of coordinates which is moving with a constant velocity, i.e., it is not accelerating.

Two events which are simultaneous in one inertial frame are not necessarily simultaneous in other inertial frames.

that there is no need to hypothesize a medium called the ether which is at rest in one preferred frame.

It now seems obvious that, if Maxwell's equations had to retain their validity in all inertial frames, then our Newtonian notions of space and time need to be modified. However, while many of the best physicists of the time were aware of this inconsistency between Maxwell and Newton, it took almost two decades after the Michelson-Morley experiment and the genius of Einstein to come up with the right modification to Newtonian mechanics.

3. Laws of Transformation of Space and Time

Einstein was a great believer in simplicity and universality. He therefore made the bold hypothesis that *all* the laws of physics, not just Maxwell's equations, must be the same in all inertial frames. In other words, there is no special frame in nature which can be considered to be at absolute rest. Einstein soon realized that this idea of 'democracy' between all inertial frames meant that our earlier notion of time needed to be modified. In typical Einsteinian style, his first paper on special relativity gets to the crux of the matter immediately. He therefore begins with a discussion of the idea of simultaneity, namely, *are two events which are simultaneous in one inertial frame necessarily simultaneous in other inertial frames also?*

Einstein showed that the constancy of the speed of light implies that simultaneity is not an absolute concept. Hence the difference in the time coordinates of two events depends on the inertial frame in which it is measured! This was a radical departure from the earlier notion of absolute time. Indeed, this concept of 'relativity of simultaneity' underlies most of the puzzles and paradoxes of relativity. Once we understand this idea, we can modify our intuition to include the strange world of relativity!

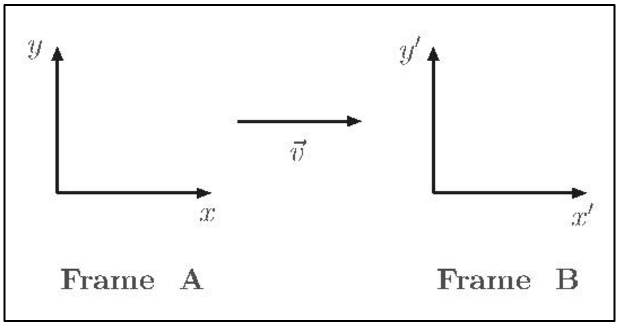

Figure 1. A picture of two inertial frames A and B. Frame B is moving with a velocity $\vec{v} = v\hat{x}$ with respect to frame A.

To establish this more clearly, let us first see how we transform from one inertial frame to another in Newtonian mechanics. Newton teaches us that if an inertial frame B is moving with respect to another inertial frame A with a velocity $\vec{v} = v\hat{x}$ (see *Figure* 1), and a point in spacetime (called an 'event') has the coordinates (x, y, z, t) in frame A and (x', y', z', t') in frame B, then these are related as:

$$\begin{aligned} x' &= x - vt \ , \\ y' &= y \ , \\ z' &= z \ , \\ \text{and} \quad t' &= t \ . \end{aligned} \qquad (1)$$

These relations are called Galilean transformations. From this, we see that the time of occurrence of an event is the same in all inertial frames. A more precise way of stating this is that the time interval between two events is *invariant*.

To retain the validity of Maxwell's equations in all frames, Einstein then showed that the Galilean transformation laws need to be changed. He introduced two postulates:

1. *The laws of physics take the same form in all inertial frames.*

2. *The speed of light in vacuum is the same in all inertial frames*[3].

[3] Actually, the second postulate follows from the first since the laws of physics include Maxwell's equations of electromagnetism which contain the speed of light.

To derive the correct transformation laws, let the two inertial frames coincide at $t = 0$. From the above postulates, it follows that if a body emits light at a space-time point with coordinates $(0, 0, 0, 0)$ in both frames, the wave front at later times must be given by the surface of a sphere which is expanding with speed c in both frames. In other words, the wave front must satisfy

$$\begin{aligned} x^2 + y^2 + z^2 &= c^2 t^2 \\ \text{and} \quad x'^2 + y'^2 + z'^2 &= c^2 t'^2 \end{aligned} \tag{2}$$

in frames A and B, respectively. Using these equations and the fact that frame B is moving with respect to frame A with speed v in the $\hat{x}$ direction, Einstein derived the relations

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}} , \\ y' &= y , \\ z' &= z , \\ \text{and} \quad t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} . \end{aligned} \tag{3}$$

These transformation laws had already been discovered by H A Lorentz in 1904 (and even earlier by W Voigt in 1887), except that he had not understood the significance of the time t'; he thought of it as some kind of a 'local' time which is needed in order to make sense of Maxwell's equations in frame B. Lorentz continued to believe in the notion of a preferred frame and an absolute time t. It was Einstein who first understood the true significance of the transformation of the time coordinate. However, the relations in equation (3) are still called the Lorentz transformations since he had discovered them one year earlier.

Under Galilean transformations, we saw that two observers agree on the time interval between two events.

Under Lorentz transformation, we see that the time interval is not invariant. The curious reader might therefore wonder what, if anything, do the two observers agree on? The answer lies in equation (2). The quantity $(x^2 + y^2 + z^2 - c^2t^2)$ is obviously the same for both observers. This is called the spacetime interval, and is a *Lorentz invariant* since it remains the same under Lorentz transformations. We will see more about Lorentz invariants later. One other point to note is that the Galilean transformation can be thought of as a limiting case of the Lorentz transformation, in the limit $c \to \infty$.

Length contraction is a consequence of the fact that simultaneity in time is not an absolute concept.

4. Consequences of the Lorentz Transformation Laws

Several results can be derived from the Lorentz transformation laws. The reader is encouraged to derive the first two results discussed below (length contraction and time dilation), which only involve some simple algebraic manipulations.

Fitzgerald-Lorentz Length Contraction: If an object of length L (as measured in its 'rest' frame B) is moving with velocity $\vec{v}$ in the direction of its length with respect to a frame A, its length as measured in frame A is given by

$$L_A = L\sqrt{1 - v^2/c^2}\ . \qquad (4)$$

This happens because length is defined to be the difference between the spatial coordinates of the two ends of an object measured at the *same time*. In other words, if we denote the two ends of the object as 1 and 2, then in frame B, $t'_2 = t'_1$ implies $x'_2 - x'_1 = L$. However, in frame A, $t_2 = t_1$ implies $x_2 - x_1 = L\sqrt{1 - v^2/c^2}$. Thus length contraction is a consequence of the fact that simultaneity in time is not an absolute concept.

Time Dilation: If in a frame B, the time difference between two events occurring at the same point in space is Δt, in a different frame A which is moving with a velocity v with respect to B, the time difference between the same two events is given by

$$\Delta t_A = \frac{\Delta t}{\sqrt{1 - v^2/c^2}} . \quad (5)$$

There is a simple way of deriving this important result which does not require a complete knowledge of the Lorentz transformation laws given earlier.

Suppose that frame B is moving along the x-axis with speed v with respect to frame A. In frame B, imagine a light wave that starts at a point O, goes up along the $+\hat{y}$ direction for a distance L, gets reflected back along the $-\hat{y}$ direction, and returns to the point O (see *Figure* 2). The two spacetime events are thus defined as (1) light pulse being emitted from O, and (2) light pulse being received at O. Since the speed of light is c, the time elapsed between the two events in frame B is $2L/c$.

Now, let us consider the two events from the point of view of frame A. First, note that a coordinate which is measured in a direction perpendicular to the relative motion of two frames has the same value in both frames. Thus the distance L in the $\hat{y}$ direction is the same in frames A and B. However, during the time interval between the two events, the point O has moved

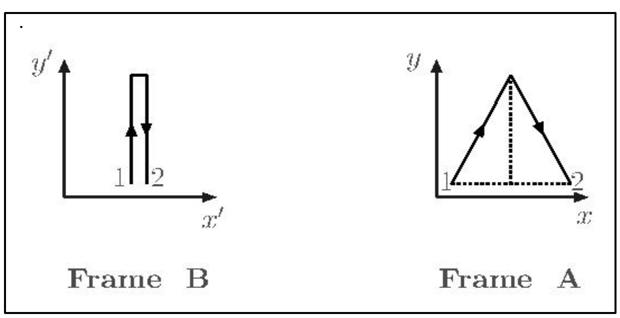

Figure 2. The path travelled by a light wave as seen in two inertial frames A and B. Frame B is moving with a velocity $\vec{v} = v\hat{x}$ with respect to frame A.

some distance to the right relative to A. Clearly, the light wave has traversed a greater distance (along the triangle) in frame A. If the speed of light with respect to A is also to be c, the time interval has to be *longer*.

To see this quantitatively, let us denote the time interval between the above two events in frame A as $\gamma 2L/c$, where γ is the time-dilation factor that we are going to derive. The point O moves a distance $(\gamma 2L/c)v$ between the two events. The time interval between the two events is therefore given by twice the time needed for light to travel the hypotenuse of a right-angled triangle; by the Pythagoras theorem, the hypotenuse has a length given by $L\sqrt{1+\gamma^2 v^2/c^2}$. Hence the time interval in frame A is $(2L/c)\sqrt{1+\gamma^2 v^2/c^2}$ which must be equal to $\gamma 2L/c$. This immediately leads to the relation

$$\gamma = \frac{1}{\sqrt{1-v^2/c^2}} \,. \tag{6}$$

The time interval viewed from a moving frame is always longer than the corresponding time interval in the rest frame.

It is important to note that time dilation is a completely symmetric phenomenon. In other words, if we repeated the above experiment with the light beam going up and down vertically in frame A, then it is an observer in frame B who will see a triangular path and hence a longer time interval. The reason for this symmetry is that all (uniform) motion is relative, and it is incorrect to say that A is moving or that B is moving; they are just moving *relative* to each other. What is true is that the time interval (or the rate of a clock) viewed from a moving frame is always longer than the corresponding time interval in the rest frame. It is in this sense that the phrase 'moving clocks go slower' is used. The time interval in the rest frame is called the *proper time*.

Velocity Addition Law: If frame B is moving with respect to frame A with velocity v_1, and frame C is moving with respect to frame A with velocity v_2 in the *opposite* *c* is the 'speed limit' of nature which no object can ever exceed.

direction, then the relative velocity between frames B and C is given by

$$v_3 = \frac{v_1 + v_2}{1 + v_1 v_2 / c^2} \ . \tag{7}$$

Einstein actually derived a more general relation in which frames B and C could move at arbitrary angles with respect to frame A.

An interesting consequence of this relation is that if both v_1 and v_2 are less than c in magnitude, then so is v_3. Thus, one cannot exceed the speed c no matter how one 'adds' up the velocities of several frames. The speed c is the 'speed limit' of nature which no object can ever exceed. Furthermore, if one of the velocities is the speed of light c, then the sum is also c. Thus, the speed of light viewed from any frame is c, which is indeed one of the postulates on which the above equation is based. Finally, note that if v_1 and v_2 are both much smaller than c, then we recover the 'intuitive' law of addition of velocities, $v_3 = v_1 + v_2$. Our intuition is based on experiences with speeds that are a small fraction of the speed of light, therefore it is not surprising that we did not realize the need for the correct law until Maxwell and then Einstein came along!

Doppler Shift: If light waves moving in some direction have a frequency ν in a frame A, then they have a frequency

$$\nu' = \nu \sqrt{\frac{1 - v/c}{1 + v/c}} \tag{8}$$

in a frame which is moving in the *same* direction with velocity v, and a frequency

$$\nu' = \nu \sqrt{\frac{1 + v/c}{1 - v/c}} \tag{9}$$

in a frame which is moving in the *opposite* direction with velocity v. Note again that if $v/c \ll 1$, one recovers the more familiar result for Doppler shift $\nu'/\nu = 1 \pm v/c$ (depending on the relative direction of the light waves and the frame of the observer).

The paper in which Einstein derived the famous relation $E=mc^2$ is remarkably short!

Using arguments from the classical theory of electromagnetism, Einstein also showed that the energy of the light waves transforms in the same way as the frequency. This is now obvious from the quantum theory of photons; the energy of a photon is proportional to the frequency of the corresponding light wave, the constant of proportionality being given by Planck's constant h.

5. The now Famous Relation $E = mc^2$

The second paper of Einstein (in which he derived a relation between the energy and the mass of an object) was remarkably short! His argument went as follows. Let an object which is at rest in a frame A simultaneously emit two light waves with the same energy $E/2$ in opposite directions. Since the two waves carry equal but opposite momenta, the object remains at rest, but its energy decreases by E.

By the Doppler shift argument given above, in a frame B which is moving at velocity v in one of those directions, the object will appear to lose an energy equal to

$$\frac{E}{2}\sqrt{\frac{1-v/c}{1+v/c}} + \frac{E}{2}\sqrt{\frac{1+v/c}{1-v/c}} = \frac{E}{\sqrt{1-v^2/c^2}} \ . \quad (10)$$

The difference in energy loss as viewed from the two frames must therefore appear as a difference in kinetic energy seen by frame B. Hence, if v/c is very small, in frame B the object loses an amount of kinetic energy

A quantity which has the same value in all inertial frames is called a Lorentz invariant.

given by

$$\frac{E}{\sqrt{1-v^2/c^2}} - E \simeq \frac{1}{2} \times \frac{E}{c^2} \times v^2 . \quad (11)$$

Since the kinetic energy of an object with mass M moving with speed v is given by $(1/2)Mv^2$ (for $v/c \ll 1$), this means that the object has lost an amount of mass given by E/c^2. In other words, a loss in energy of E is equivalent to a loss in mass of E/c^2. This implies an equivalence between the mass and energy content of any object.

Transformation Laws of Energy and Momentum: It turns out that not only the spacetime coordinates, but also some other sets of four quantities transform the same way as given by equation (3). For instance, if E and (p_x, p_y, p_z) denote the energy and three components of the momentum of a particle, then E/c^2 and (p_x, p_y, p_z) transform in exactly the same way as t and (x, y, z). This means that the quantity $E^2/c^2 - p_x^2 - p_y^2 - p_z^2$ has the same value in all inertial frames, i.e. it is a Lorentz invariant (just like the quantity $x^2+y^2+z^2-c^2t^2$ discussed earlier). It turns out that for a particle of mass M, this quantity is equal to M^2c^2. Thus the energy and momentum of a particle are related as

$$E = \sqrt{M^2c^4 + p_x^2c^2 + p_y^2c^2 + p_z^2c^2} . \quad (12)$$

If a particle has zero momentum (the frame in which this is true is called the rest frame of that particle), then we recover the relation $E = Mc^2$. If a particle has a non-zero momentum but the components of the momentum are all much smaller than Mc, we can expand the right hand side of (12) to obtain

$$E = Mc^2 + \frac{p_x^2 + p_y^2 + p_z^2}{2M} + \cdots . \quad (13)$$

The first term is the rest-mass energy while the second term is the form of the kinetic energy that we are familiar with in non-relativistic classical mechanics.

The muon appears to live seven times longer in the frame of the Earth.

Finally, we can use the Lorentz transformations in equation (3) to calculate the energy of a particle in different inertial frames. We know that the energy and momentum of a particle are given by Mc^2 and zero respectively in its rest frame. We then find that in a frame which is moving with a speed v with respect to its rest frame, its energy is given by

$$E = \frac{Mc^2}{\sqrt{1 - v^2/c^2}} \,. \tag{14}$$

6. Experimental Verification of Einstein's Ideas

The special theory of relativity has been experimentally tested in a variety of ways over the years. We present below some of the more striking consequences of this theory.

Muon Decay: There is an unstable particle called the muon which decays into other particles with a time constant of $\tau = 2$ microseconds. These particles are created when cosmic rays hit the upper atmosphere, at a distance of tens of kilometers above the Earth's surface. The particles are created at speeds of up to 0.99 c, but even at such high speeds, one would expect that they travel only about 600 meters (in 2 microseconds) before decaying. However, some muons are found near the Earth's surface!

The explanation for this observation comes from the phenomenon of time dilation. The lifetime of 2 microseconds for the muon is true in its rest frame. In the Earth-bound frame, the muon is moving at a speed of 0.99 c, hence its lifetime is given by $\tau/\sqrt{1 - v^2/c^2} \simeq 7\,\tau \simeq 14$ microseconds. The average distance that they travel in one lifetime is about 4 kilometers, and some of them do reach the Earth's surface.

Fission and Fusion: In general, the mass of an atom is *less* than the sum of the masses of the neutrons, protons and electrons contained in it. The mass difference is due to the binding energy; when particles come together to form a bound state, they lose the binding energy E_B and therefore a mass given by E_B/c^2. Most of the binding energy of an atom is due to the nuclear forces which bind neutrons and protons together to form the nucleus; the binding energy of the electrons to the nucleus is about a million times smaller. Roughly speaking, this means that chemical interactions (involving electrons) are a million times weaker than nuclear interactions.

If we look at the binding energy of a nucleus and divide it by the total number of nucleons[4], we get an average value of the binding energy called the binding energy per nucleon. This is obviously different for different nuclei, but we find that it is largest for an isotope of iron, namely ${}^{56}_{26}\mathrm{Fe}$. In this notation, ${}^{56}_{26}\mathrm{Fe}$ means that the nucleus of the iron atom (Fe) contains 26 protons and $56 - 26 = 30$ neutrons. The binding energy per nucleon for this nucleus is 8.8 million electron volts (MeV), and becomes progressively smaller both for larger and smaller nuclei[5]. For example, the binding energy per nucleon for the alpha particle (${}^{4}_{2}\mathrm{He}$) is 7.1 MeV. Thus, heavy nuclei are comparatively less stable and can be split into lighter nuclei that are more stable. This process is called fission. Similarly, very light nuclei can be joined together to form more stable heavier nuclei. This process is called fusion. In both processes, the difference in binding energy is released in the form of kinetic energy of the daughter particles.

An example of a fission process is the decay of uranium. When a neutron strikes a uranium nucleus, the latter can decay into a cesium nucleus and a rubidium nucleus

[4] Neutrons and protons are collectively called nucleons because they form the constituents of all nuclei.

[5] In conventional units, 1 MeV $\simeq 1.602 \times 10^{-13}$ Joules.

emitting three neutrons and some energy,

$$ {}^{1}_{0}\mathrm{n} + {}^{235}_{92}\mathrm{U} \rightarrow {}^{140}_{55}\mathrm{Cs} + {}^{93}_{37}\mathrm{Rb} + 3\,{}^{1}_{0}\mathrm{n} + 200 \text{ MeV} . \quad (15) $$

Since the process starts with one neutron and results in three neutrons being produced, this can have a multiplicative effect in that each of the daughter neutrons can cause further fission of a uranium nucleus, setting off what is called a chain reaction. This is the reaction used in most nuclear power plants today. The fission of 1 kg of uranium releases 18.7 million kilowatt-hours as heat.

The energy released in fission and fusion is due to a mismatch in the masses of the initial and final particles.

An example of a fusion process is the reaction by which two nuclei of a heavy isotope of hydrogen (called the deuteron) join to form a helium nucleus, thereby emitting one neutron and some energy,

$$ {}^{2}_{1}\mathrm{H} + {}^{2}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + {}^{1}_{0}\mathrm{n} + 3.2 \text{ MeV}. \quad (16) $$

This is one of the main processes that is responsible for the production of energy in most stars including the sun. Some of this energy eventually reaches the Earth and makes life possible here!

High-energy Physics: In high-energy collisions, it is common for particles to appear or disappear completely! This seems to violate what we have learnt in school that matter cannot be created or destroyed. The explanation comes from the energy-mass relation, which allows a particle's mass to be *entirely* converted to other forms of energy and *vice versa*[6]. Of course, a quick calculation will show that you need large amounts of energy to create even a light particle such as the electron, hence these processes are observed only in *high-energy* experiments. High-energy physicists therefore find it convenient to express the masses of particles in terms of their equivalent energy. For instance, the mass of an electron is about

[6] The entire conversion of mass to energy is in contrast to nuclear fission and fusion discussed above, where only about 0.1% of the mass is converted to other forms of energy.

The existence of antiparticles is an unexpected consequence of the special theory of relativity.

$0.511\ \text{MeV}/c^2$, (about 9.109×10^{-31} kg in conventional terms), and the mass of a proton is about 938 MeV/c^2 (about 1.673×10^{-27} kg).

A modern particle accelerator such as the Tevatron in Fermilab near Chicago can collide protons and antiprotons with a total energy of 2 million MeV. In fact, the name Tevatron comes from tera electron volt (TeV), which is equal to a million MeV. With such an energy, one can create a particle which is two thousand times more massive than a proton. In 2007, a more powerful accelerator called the LHC (Large Hadron Collider) will begin working in Geneva; this will collide protons with protons with a total energy of 14 million MeV.

Existence of Antiparticles: This is a less-fre- quently discussed consequence of the special theory of relativity, one which Einstein did not anticipate because antiparticles were not known in 1905. Given a particle with mass M and charge Q, its antiparticle is an object with the same mass M but the opposite charge Q. Some neutral particles like the neutron also have antiparticles; these differ from the particle in some other ways.

To see why the existence of antiparticles follows necessarily from the special theory of relativity, consider the sequence of events shown in *Figure* 3. In the figure, time increases in the vertical direction while spatial coordinates (say, the x coordinate) changes in the horizontal

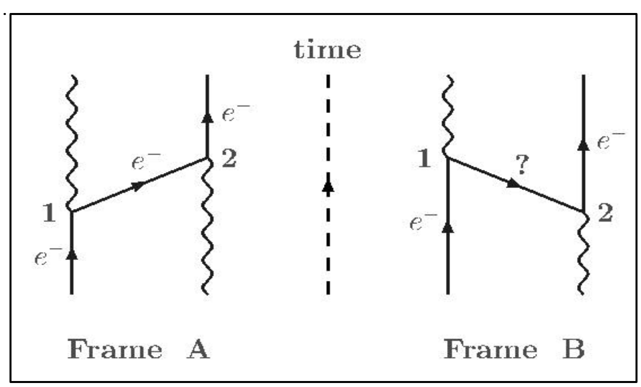

Figure 3. A process in which an electron (e⁻) emits and absorbs a photon at two different spacetime points, as seen in two inertial frames A and B. The identity of the mysterious particle marked by a '?' in the middle of frame B will be revealed in the next figure.

Box 1

A simple numerical example will illustrate how the time ordering of two events is not an absolute notion. In an inertial frame A, let us suppose that event 1 occurs 10^{-8} seconds before event 2, and that they are separated by a distance of 90 centimeters in the $\hat{x}$ direction. Now consider a frame B which is moving with respect to frame A with a velocity of 0.99 c in the $\hat{x}$ direction. Using the Lorentz transformations in equation (3), we find that in this frame, event 1 occurs 10^{-8} seconds after event 2. This is a very small time difference by human standards; the smallness is due to the high speed of light which takes only 3×10^{-9} seconds to travel a distance of 90 centimeters. Also, whenever this happens, i.e. the time sequence of two events gets reversed, one of them could not have been the cause of the other! The two events are outside each other's `sphere of influence'.

direction. An electron (denoted by e^- and shown by a solid line) emits a photon (shown by a wavy line) at the spacetime point 1 and absorbs a photon at the spacetime point 2. The photons do not play any role in the argument given below; however, they are shown to indicate the locations of the points 1 and 2 where the electron undergoes a change in energy and momentum due to the emission or absorption of a photon.

Let us assume that in frame A, event 2 occurs after event 1, but in frame B, event 2 occurs *before* event 1. This is possible in relativity because the time ordering of two events is not an absolute concept – one event can be in the past of another event in one frame and in its future in another frame (see *Box* 1). Now the question is: what are the various particles that an observer will see at different times in the two frames? (We will ignore the photons in this discussion). In frame A, we clearly see an electron before event 1, an electron between events 1 and 2, and an electron after event 2. In frame B, we again see only one electron before event 2, and only one electron after event 1.

Question: What particles are observed in frame B between events 2 and 1? In particular, what is the mysterious particle marked with a '?' in the middle of frame B in *Figure* 3?

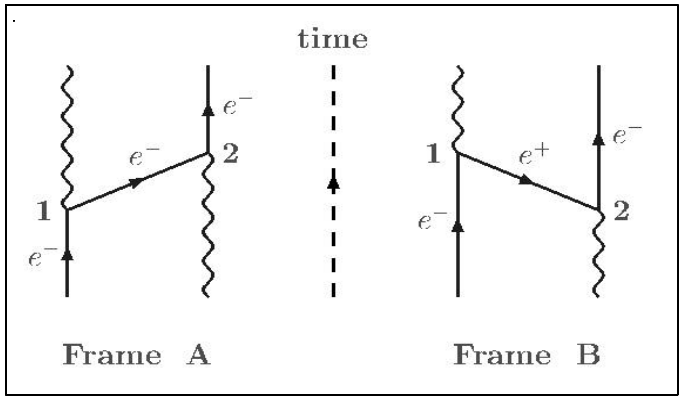

Figure 4. The same process as shown in the previous figure. The particle in the middle of frame B has to be the antiparticle of the electron (e⁺) in order to conserve the total charge at all times.

Answer: In frame B, we have three particles of the same mass. This is because, as we have discussed earlier, the mass of a particle is the same in all inertial frames. Since the particle in the middle has the mass of an electron in frame A, it must have the mass of an electron in frame B also. The two particles in the beginning and end are clearly electrons. The third particle (in the middle) cannot be an electron because the total charge must remain the same at all times in any frame. This is required by the law of conservation of charge, which can never be violated under any circumstances. Hence, the charge of the third particle must be opposite to that of an electron, so that the total charge at all times remains equal to the charge of one electron. Thus, the third particle has the same mass as the electron but has the opposite charge. This is the antiparticle of the electron, and it is called the positron (denoted by e^+). The complete picture is shown in *Figure* 4. It is important to note that, in frame B, the positron is created at point 2, and then travels to point 1 where it is annihilated by the electron.

7. Final Comments

The most important lesson from the special theory of relativity is that simultaneity is relative. As mentioned earlier, this concept underlies many of the apparent paradoxes in relativity. The phrase 'at the same time' is an observer-dependent term, and most paradoxes will be solved if the use of this phrase is removed. The reader is encouraged to have a fresh look at the puzzles discussed in many books on relativity with this idea in mind.

The 'general theory of relativity' is probably the greatest intellectual creation to come from a single person.

Since the speeds of most objects that we see around us are much smaller than c, one might think that the differences between non-relativistic mechanics and relativistic mechanics would be too small to be of any practical importance in our daily life. While this is generally true, there are some situations where it becomes essential to take relativistic corrections into account. An interesting example is provided by the satellites that implement the global positioning system. These satellites carry atomic clocks which can measure very precisely the time differences between messages sent out and received by different satellites and ground stations. In order to use these time differences to calculate the position of an object on the Earth's surface with sufficient accuracy (of the order of centimeters), it turns out to be essential to use the special theory of relativity. The curious reader may look at [8] for a detailed discussion of this point.

We should also emphasize that the special theory of relativity is not so much a theory as a guiding principle for physics. It tells us how to formulate the laws of physics by directing us to write them in a form that will be the same for all inertial observers. Such laws are said to be relativistically *covariant*. It also informs us that space and time can be transformed into each other, and the parameter that expresses this relation between space and time is c. Thus, c will appear naturally in any relativistically covariant theory. It is just a historical accident that the first relativistic theory we discovered was the Maxwell theory of electromagnetism, and as a result c is called the speed of light. If we had instead discovered a relativistic theory of gravitation to supersede Newton's law, we might have equally well called c as the speed of gravity!

It appears quite natural to try and extend the principle of special relativity to a more general principle that the laws of physics should be the same in *all* frames, both inertial and accelerated. This is indeed what Einstein did 10 years after developing the special theory. The extended theory is called the 'general theory of relativity', and is probably the greatest intellectual creation to come from a single person. It is a profoundly beautiful theory, although it only includes gravitation in its description. The general theory led to an even greater revolution in physics by giving rise to the idea of curved spacetime, nonlinear equations where force and inertial manifestations are combined, the idea of black holes, and so much more, but that is a story for another day.

Address for Correspondence
Vasant Natarajan
Department of Physics
Indian Institute of Science
Bangalore 560 012, India.
Email:
vasant@physics.iisc.ernet.in
http://physics.iisc.ernet.in/~vasant

Diptiman Sen
Centre for High Energy Physics
Indian Institute of Science
Bangalore 560 012, India.
Email:
diptiman@cts.iisc.ernet.in

Suggested Reading

[1] George Gamow, *Mr Tompkins in Paperback*, Cambridge University Press, Cambridge, 1995.
[2] H A Lorentz, A Einstein, H Minkowski and H Weyl, *The Principle of Relativity*, Dover Publications, London, 1952.
[3] A Pais, *'Subtle is the Lord...' The Science and the Life of Albert Einstein*, Oxford University Press, Oxford, 1982.
[4] Albert Einstein, *The Meaning of Relativity*, Fifth Edition, Princeton University Press, Princeton, 1956.
[5] Robert Resnick, *Introduction to Special Relativity*, Wiley, New York, 1968.
[6] Supurna Sinha, Einstein and the Special Theory of Relativity, *Resonance*, Vol. 5, No. 3, p.6, 2000.
[7] http://www.aip.org/history/einstein/
[8] Neil Gershenfeld, *The Physics of Information Technology*, Cambridge University Press, Cambridge, 2000.

"sometimes ask myself how it came about that I was the one to develop the theory of relativity. The reason, I think, is that a normal adult never stops to think about problems of space and time. These are things which he has thought about as a child. But my intellectual development was retarded, as a result of which I began to wonder about space and time only when I had already grown up."

Einstein on relativity