Introduction

On the cosmic scale, gravitation dominates the universe. Nuclear and electromagnetic forces account for the detailed processes that allow stars to shine and astronomers to see them. But it is gravitation that shapes the universe, determining the geometry of space and time and thus the large-scale distribution of galaxies. Providing insight into gravitation – its effects, its nature and its causes – is therefore rightly seen as one of the most important goals of physics and astronomy.

Figure 1 Albert Einstein (1879–1955) depicted during the time that he worked at the Patent Office in Bern. While there, he published a series of papers relating to special relativity, quantum physics and statistical mechanics. He was awarded the Nobel Prize for Physics in 1921, mainly for his work on the photoelectric effect.

Through more than a thousand years of human history the common explanation of gravitation was based on the Aristotelian belief that objects had a natural place in an Earth-centred universe that they would seek out if free to do so. For about two and a half centuries the Newtonian idea of gravity as a force held sway. Then, in the twentieth century, came Einstein's conception of gravity as a manifestation of spacetime curvature. It is this latter view that is the main concern of this book.

The story of Einsteinian gravitation begins with a failure. Einstein's theory of special relativity, published in 1905 while he was working as a clerk in the Swiss Patent Office in Bern, marked an enormous step forward in theoretical physics and soon brought him academic recognition and personal fame. However, it also showed that the Newtonian idea of a gravitational force was inconsistent with the relativistic approach and that a new theory of gravitation was required. Ten years later, Einstein's general theory of relativity met that need, highlighting the important role of geometry in accounting for gravitational phenomena and leading on to concepts such as black holes and gravitational waves. Within a year and a half of its completion, the new theory was providing the basis for a novel approach to cosmology – the science of the universe – that would soon have to take account of the astronomy of galaxies and the physics of cosmic expansion. The change in thinking demanded by relativity was radical and profound. Its mastery is one of the great challenges and greatest delights of any serious study of physical science.

This book begins with two chapters devoted to special relativity. These are followed by a mainly mathematical chapter that provides the background in geometry that is needed to appreciate Einstein's subsequent development of the theory. Chapter 4 examines the basic principles and assumptions of general relativity – Einstein's theory of gravity – while Chapters 5 and 6 apply the theory to an isolated spherical body and then extend that analysis to non-rotating and rotating black holes. Chapter 7 concerns the testing of general relativity, including the use of astronomical observations and gravitational waves. Finally, Chapter 8 examines modern relativistic cosmology, setting the scene for further and ongoing studies of observational cosmology.

The text before you is the result of a collaborative effort involving a team of authors and editors working as part of the broader effort to produce the Open University course S383 *The Relativistic Universe*. Details of the team's membership and responsibilities are listed elsewhere but it is appropriate to acknowledge here the particular contributions of Jim Hague regarding Chapters 1 and 2, Derek Capper concerning Chapters 3, 4 and 7, and Aiden Droogan in relation to Chapters 5, 6 and 8. Robert Lambourne was responsible for planning and producing the final unified text which benefited greatly from the input of the S383 Course Team Chair, Andrew Norton, and the attention of production editor

Peter Twomey. The whole team drew heavily on the work and wisdom of an earlier Open University Course Team that was responsible for the production of the course S357 *Space, Time and Cosmology*.

A major aim for this book is to allow upper-level undergraduate students to develop the skills and confidence needed to pursue the independent study of the many more comprehensive texts that are now available to students of relativity, gravitation and cosmology. To facilitate this the current text has largely adopted the notation used in the outstanding book by Hobson et al.

General Relativity : An Introduction for Physicists, M. P. Hobson, G. Efstathiou and A. N. Lasenby, Cambridge University Press, 2006.

Other books that provide valuable further reading are (roughly in order of increasing mathematical demand):

An Introduction to Modern Cosmology, A. Liddle, Wiley, 1999.
Relativity, Gravitation and Cosmology : A Basic Introduction, T-P. Cheng, Oxford University Press: 2005.
Introducing Einstein's Relativity, R. d'Inverno, Oxford University Press, 1992.
Relativity : Special, General and Cosmological, W. Rindler, Oxford University Press, 2001.
Cosmology, S. Weinberg, Cambridge University Press, 2008.

Two useful sources of reprints of original papers of historical significance are:

The Principle of Relativity, A. Einstein et al., Dover, New York, 1952.
Cosmological Constants, edited by J. Bernstein and G. Feinberg, Columbia University Press, 1986.

Those wishing to undertake background reading in astronomy, physics and mathematics to support their study of this book or of any of the others listed above might find the following particularly helpful:

An Introduction to Galaxies and Cosmology, edited by M. H. Jones and R. J. A. Lambourne, Cambridge University Press, 2003.
The seven volumes in the series
The Physical World, edited by R. J. A. Lambourne, A. J. Norton et al., Institute of Physics Publishing, 2000.
(Go to www.physicalworld.org for further details.)
The paired volumes
Basic Mathematics for the Physical Sciences, edited by R. J. A. Lambourne and M. H. Tinker, Wiley, 2000.
Further Mathematics for the Physical Sciences, edited by M. H. Tinker and R. J. A. Lambourne, Wiley, 2000.

Numbered exercises appear throughout this book. Complete solutions to these exercises can be found at the back of the book. There are a number of lengthy worked examples; these are highlighted by a blue background. There are also several shorter in-text questions that are immediately followed by their answers. These may be treated as questions or examples. The questions are indented and indicated by a filled circle; their answers are also indented and shown by an open circle.

Chapter 1 Special relativity and spacetime

Introduction

In two seminal papers in 1861 and 1864, and in his treatise of 1873, James Clerk Maxwell (Figure 1.1), Scottish physicist and genius, wrote down his revolutionary unified theory of electricity and magnetism, a theory that is now summarized in the equations that bear his name. One of the deep results of the theory introduced by Maxwell was the prediction that wave-like excitations of combined electric and magnetic fields would travel through a vacuum with the same speed as light. It was soon widely accepted that light itself was an electromagnetic disturbance propagating through space, thus unifying electricity and magnetism with optics.

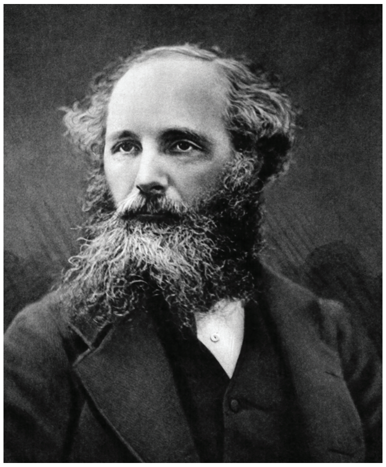

Figure 1.1 James Clerk Maxwell (1831–1879) developed a theory of electromagnetism that was already compatible with special relativity theory several decades before Einstein and others developed the theory. He is also famous for major contributions to statistical physics and the invention of colour photography.

The fundamental work of Maxwell opened the way for an understanding of the universe at a much deeper level. Maxwell himself, in common with many scientists of the nineteenth century, believed in an all-pervading medium called the **ether**, through which electromagnetic disturbances travelled, just as ocean waves travelled through water. Maxwell's theory predicted that light travels with the same speed in all directions, so it was generally assumed that the theory predicted the results of measurements made using equipment that was at rest with respect to the ether. Since the Earth was expected to move through the ether as it orbited the Sun, measurements made in terrestrial laboratories were expected to show that light actually travelled with different speeds in different directions, allowing the speed of the Earth's movement through the ether to be determined. However, experiments, most notably by A. A. Michelson and E. W. Morley in 1887, failed to detect any variations in the measured speed of light. This led some to suspect that measurements of the speed of light in a vacuum would always yield the same result irrespective of the motion of the measuring equipment. Explaining how this could be the case was a major challenge that prompted ingenious proposals from mathematicians and physicists such as Henri Poincaré, George Fitzgerald and Hendrik Lorentz. However, it was the young Albert Einstein who first put forward a coherent and comprehensive solution in his 1905 paper 'On the electrodynamics of moving bodies', which introduced the **special theory of relativity**. With the benefit of hindsight, we now realize that Maxwell had unintentionally formulated the first major theory that was consistent with special relativity, a revolutionary new way of thinking about space and time.

This chapter reviews the implications of special relativity theory for the understanding of space and time. The narrative covers the fundamentals of the theory, concentrating on some of the major differences between our intuition about space and time and the predictions of special relativity. By the end of this chapter, you should have a broad conceptual understanding of special relativity, and be able to derive its basic equations, *the Lorentz transformations*, from the postulates of special relativity. You will understand how to use *events* and *intervals* to describe properties of space and time far from gravitating bodies. You will also have been introduced to Minkowski spacetime, a four-dimensional fusion of space and time that provides the natural setting for discussions of special relativity.

1.1 Basic concepts of special relativity

1.1.1 Events, frames of reference and observers

When dealing with special relativity it is important to use language very precisely in order to avoid confusion and error. Fundamental to the precise description of physical phenomena is the concept of an *event*, the spacetime analogue of a point in space or an instant in time.

Events

An **event** is an instantaneous occurrence at a specific point in space.

An exploding firecracker or a small light that flashes once are good approximations to events, since each happens at a definite time and at a definite position.

To know when and where an event happened, we need to assign some coordinates to it: a time coordinate t and an ordered set of spatial coordinates such as the *Cartesian coordinates* (x, y, z), though we might equally well use *spherical coordinates* (r, θ, ϕ) or any other suitable set. The important point is that we should be able to assign a unique set of clearly defined coordinates to any event. This leads us to our second important concept, a *frame of reference*.

Frames of reference

A **frame of reference** is a system for assigning coordinates to events. It consists of a system of synchronized clocks that allows a unique value of the time to be assigned to any event, and a system of spatial coordinates that allows a unique position to be assigned to any event.

In much of what follows we shall make use of a Cartesian coordinate system with axes labelled x, y and z. The precise specification of such a system involves selecting an origin and specifying the orientation of the three orthogonal axes that meet at the origin. As far as the system of clocks is concerned, you can imagine that space is filled with identical synchronized clocks all ticking together (we shall need to say more about how this might be achieved later). When using a particular frame of reference, the time assigned to an event is the time shown on the clock at the site of the event when the event happens. It is particularly important to note that the time of an event is not the time at which the event is seen at some far off point — it is the time at the event itself that matters.

Reference frames are often represented by the letter S. Figure 1.2 provides what we hope is a memorable illustration of the basic idea, in this case with just two spatial dimensions. This might be called the frame $\mathrm{S}_{\mathrm{gnome}}$.

Among all the frames of reference that might be imagined, there is a class of frames that is particularly important in special relativity. This is the class of *inertial frames*. An inertial frame of reference is one in which a body that is not subject to any net force maintains a constant velocity. This is *Newton's first law of motion*, so we can say the following.

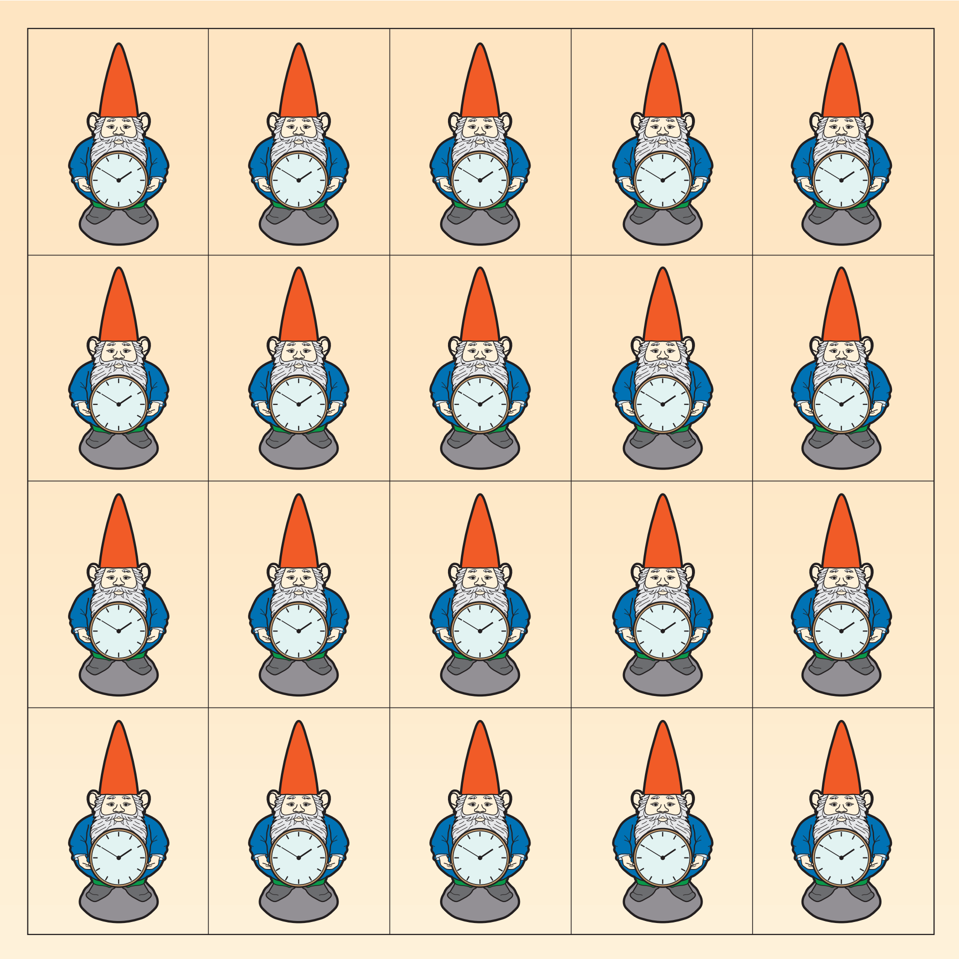

Figure 1.2 A jocular representation of a frame of reference in two spatial dimensions. Gnomes pervade all of space and time. Each gnome has a perfectly reliable clock. When an event occurs, the gnome nearest to the event communicates the time t and location (x, y) of the event to the observer.

Inertial frames of reference

An **inertial frame of reference** is a frame of reference in which Newton's first law of motion holds true.

Any frame that moves with constant velocity relative to an inertial frame will also be an inertial frame. So, if you can identify or establish one inertial frame, then you can find an infinite number of such frames, each having a constant velocity relative to any of the others. Any frame that accelerates relative to an inertial frame cannot be an inertial frame. Since rotation involves changing velocity, any frame that rotates relative to an inertial frame is also disqualified from being inertial.

One other concept is needed to complete the basic vocabulary of special relativity. This is the idea of an *observer*.

Observers

An **observer** is an individual dedicated to using a particular frame of reference for recording events.

We might speak of an observer O using frame S, or a different observer O′ (read as 'O-prime') using frame S′ (read as 'S-prime').

Though you may think of an observer as a person, just like you or me, at rest in their chosen frame of reference, it is important to realize that an observer's location is of no importance for reporting the coordinates of events in special relativity. The position that an observer assigns to an event is the place where it happened. The time that an observer assigns is the time that would be shown on a clock at the site of the event when the event actually happened, and where the clock concerned is part of the network of synchronized clocks always used in that observer's frame of reference. An observer might see the explosion of a distant star tonight, but would report the time of the explosion as the time long ago when the explosion actually occurred, not the time at which the light from the explosion reached the observer's location. To this extent, 'seeing' and 'observing' are very different processes. It is best to avoid phrases such as 'an observer sees . . . ' unless that is what you really mean. An observer measures and observes.

Any observer who uses an inertial frame of reference is said to be an **inertial observer**. Einstein's special theory of relativity is mainly concerned with observations made by inertial observers. That's why it's called *special* relativity — the term 'special' is used in the sense of 'restricted' or 'limited'. We shall not really get away from this limitation until we turn to general relativity in Chapter 4.

Exercise 1.1 For many purposes, a frame of reference fixed in a laboratory on the Earth provides a good approximation to an inertial frame. However, such a frame is not really an inertial frame. How might its true, non-inertial, nature be revealed experimentally, at least in principle? ■

1.1.2 The postulates of special relativity

Physicists generally treat the laws of physics as though they hold true everywhere and at all times. There is some evidence to support such an assumption, though it is recognized as a hypothesis that might fail under extreme conditions. To the extent that the assumption is true, it does not matter where or when observations are made to test the laws of physics since the time and place of a test of fundamental laws should not have any influence on its outcome.

Where and when laws are tested might not influence the outcome, but what about motion? We know that inertial and non-inertial observers will not agree about Newton's first law. But what about different inertial observers in uniform relative motion where one observer moves at constant velocity with respect to the other? A pair of inertial observers *would* agree about Newton's first law; might they also agree about other laws of physics?

It has long been thought that they would at least agree about the laws of mechanics. Even before Newton's laws were formulated, the great Italian physicist Galileo Galilei (1564–1642) pointed out that a traveller on a smoothly moving boat had exactly the same experiences as someone standing on the shore. A ball game could be played on a uniformly moving ship just as well as it could be played on shore. To the early investigators, uniform motion alone appeared to have no detectable consequences as far as the laws of mechanics were concerned. An observer shut up in a sealed box that prevented any observation of the outside world would be unable to perform any mechanics experiment that would reveal the uniform velocity of the box, even though any acceleration could be easily detected. (We are all familiar with the feeling of being pressed back in our seats when a train or car accelerates forward.) These notions provided the basis for the first theory of relativity, which is now known as **Galilean relativity** in honour of Galileo's original insight. This theory of relativity assumes that all inertial observers will agree about the laws of Newtonian mechanics.

Einstein believed that inertial observers would agree about the laws of physics quite generally, not just in mechanics. But he was not convinced that Galilean relativity was correct, which brought Newtonian mechanics into question. The only statement that he wanted to presume as a law of physics was that all inertial observers agreed about the speed of light in a vacuum. Starting from this minimal assumption, Einstein was led to a new theory of relativity that was markedly different from Galilean relativity. The new theory, the special theory of relativity, supported Maxwell's laws of electromagnetism but caused the laws of mechanics to be substantially rewritten. It also provided extraordinary new insights into space and time that will occupy us for the rest of this chapter.

Einstein based the special theory of relativity on two *postulates*, that is, two statements that he believed to be true on the basis of the physics that he knew. The first postulate is often referred to as the **principle of relativity**.

The first postulate of special relativity

The laws of physics can be written in the same form in all inertial frames.

This is a bold extension of the earlier belief that observers would agree about the laws of mechanics, but it is not at first sight exceptionally outrageous. It will, however, have profound consequences.

The second postulate is the one that gives primacy to the behaviour of light, a subject that was already known as a source of difficulty. This postulate is sometimes referred to as the **principle of the constancy of the speed of light**.

The second postulate of special relativity

The speed of light in a vacuum has the same constant value, $c = 3 \times 10^8\,\mathrm{m\,s^{-1}}$, in all inertial frames.

This postulate certainly accounts for Michelson and Morley's failure to detect any variations in the speed of light, but at first sight it still seems crazy. Our experience with everyday objects moving at speeds that are small compared with the speed of light tells us that if someone in a car that is travelling forward at speed v throws something forward at speed w relative to the car, then, according to an observer standing on the roadside, the thrown object will move forward with speed $v + w$. But the second postulate tells us that if the traveller in the car turns on a torch, effectively throwing forward some light moving at speed c relative to the car, then the roadside observer will find that the light travels at speed c, not the $v + c$ that might have been expected. Einstein realized that for this to be true, space and time must behave in previously unexpected ways.

The second postulate has another important consequence. Since all observers agree about the speed of light, it is possible to use light signals (or any other electromagnetic signal that travels at the speed of light) to ensure that the network of clocks we imagine each observer to be using is properly synchronized. We shall not go into the details of how this is done, but it is worth pointing out that if an observer sent a radar signal (which travels at the speed of light) so that it arrived at an event just as the event was happening and was immediately reflected back, then the time of the event would be midway between the times of transmission and reception of the radar signal. Similarly, the distance to the event would be given by half the round trip travel time of the signal, multiplied by the speed of light.

1.2 Coordinate transformations

A **theory of relativity** concerns the relationship between observations made by observers in relative motion. In the case of special relativity, the observers will be inertial observers in uniform relative motion, and their most fundamental observations will be the time and space coordinates of events.

For the sake of definiteness and simplicity, we shall consider two inertial observers O and O$'$ whose respective frames of reference, S and S$'$, are arranged in the following **standard configuration** (see Figure 1.3):

1. The origin of frame S$'$ moves along the x-axis of frame S, in the direction of increasing values of x, with constant velocity $\boldsymbol{V}$ as measured in S.
2. The x-, y- and z-axes of frame S are always parallel to the corresponding x'-, y'- and z'-axes of frame S$'$.
3. The event at which the origins of S and S$'$ coincide occurs at time $t = 0$ in frame S and at time $t' = 0$ in frame S$'$.

We shall make extensive use of 'standard configuration' in what follows. The arrangement does not entail any real loss of generality since any pair of inertial frames in uniform relative motion can be placed in standard configuration by choosing to reorientate the coordinate axes in an appropriate way, shifting the origin, and resetting the clocks appropriately.

In general, the observers using the frames S and S$'$ will not agree about the coordinates of an event, but since each observer is using a well-defined frame of reference, there must exist a set of equations relating the coordinates (t, x, y, z) assigned to a particular event by observer O, to the coordinates (t', x', y', z') assigned to the same event by observer O$'$. The set of equations that performs the task of relating the two sets of coordinates is called a **coordinate transformation**. This section considers first the *Galilean transformations* that provide the basis of Galilean relativity, and then the *Lorentz transformations* on which Einstein's special relativity is based.

1.2.1 The Galilean transformations

Before the introduction of special relativity, most physicists would have said that the coordinate transformation between S and S$'$ was 'obvious', and they would have written down the following **Galilean transformations**:

$$t' = t, \tag{1.1}$$

$$x' = x - Vt, \tag{1.2}$$

$$y' = y, \tag{1.3}$$

$$z' = z, \tag{1.4}$$

where $V = |\boldsymbol{V}|$ is the relative speed of S′ with respect to S.

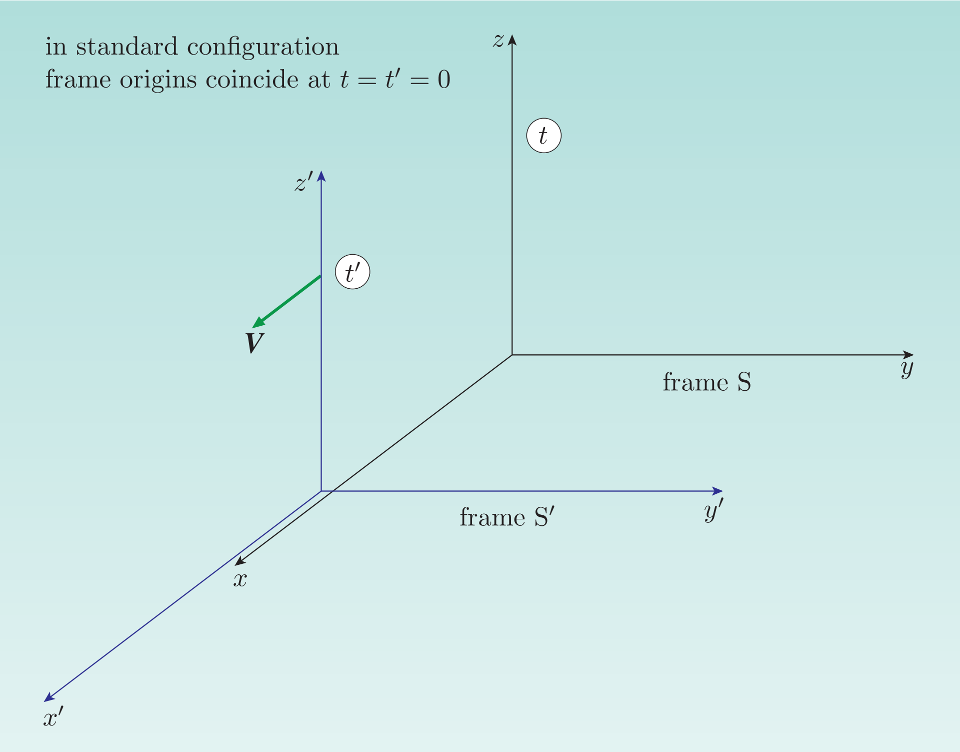

Figure 1.3 Two frames of reference in *standard configuration*. Note that the speed V is measured in frame S.

To justify this result, it might have been argued that since the observers agree about the time of the event at which the origins coincide (see point 3 in the definition of standard configuration), they must also agree about the times of all other events. Further, since at time t the origin of S′ will have travelled a distance Vt along the x-axis of frame S, it must be the case that any event that occurs at time t with position coordinate x in frame S must occur at $x' = x - Vt$ in frame S′, while the values of y and z will be unaffected by the motion. However, as Einstein realized, such an argument contains many assumptions about the behaviour of time and space, and those assumptions might not be correct. For example, Equation 1.1 implies that time is in some sense *absolute*, by which we mean that the time interval between any two events is the same for all observers. Newton certainly believed this to be the case, but without supporting evidence it was really nothing more than a plausible assumption. It was intuitively appealing, but it was fundamentally untested.

1.2.2 The Lorentz transformations

Figure 1.4 Hendrik Lorentz (1853–1928) wrote down the Lorentz transformations in 1904. He won the 1902 Nobel Prize for Physics for work on electromagnetism, and was greatly respected by Einstein.

Rather than rely on intuition and run the risk of making unjustified assumptions, Einstein chose to set out his two postulates and use them to deduce the appropriate coordinate transformation between S and S′. A derivation will be given later, but before that let's examine the result that Einstein found. The equations that he derived had already been obtained by the Dutch physicist Hendrik Lorentz (Figure 1.4) in the course of his own investigations into light and electromagnetism. For that reason, they are known as the **Lorentz transformations** even though Lorentz did not interpret or utilize them in the same way that Einstein did. Here are the equations:

$$t' = \frac{t - Vx/c^2}{\sqrt{1 - V^2/c^2}},$$
$$x' = \frac{x - Vt}{\sqrt{1 - V^2/c^2}},$$
$$y' = y,$$
$$z' = z.$$

It is clear that the Lorentz transformations are very different from the Galilean transformations. They indicate a thorough mixing together of space and time, since the t'-coordinate of an event now depends on both t and x, just as the x'-coordinate does. According to the Lorentz transformations, the two observers do not generally agree about the time of events, even though they still agree about the time at which the origins of their respective frames coincided. So, time is no longer an absolute quantity that all observers agree about. To be meaningful, statements about the time of an event must now be associated with a particular observer. Also, the extent to which the observers disagree about the x-coordinate of an event has been modified by a factor of $1/\sqrt{1 - V^2/c^2}$. In fact, this multiplicative factor is so common in special relativity that it is usually referred to as the **Lorentz factor** or **gamma factor** and is represented by the symbol $\gamma(V)$, emphasizing that its value depends on the relative speed V of the two frames. Using this factor, the Lorentz transformations can be written in the following compact form.

The Lorentz transformations

$$t' = \gamma(V)(t - Vx/c^2), \qquad (1.5)$$
$$x' = \gamma(V)(x - Vt), \qquad (1.6)$$
$$y' = y, \qquad (1.7)$$
$$z' = z, \qquad (1.8)$$

where

$$\gamma(V) = \frac{1}{\sqrt{1 - V^2/c^2}}. \qquad (1.9)$$

Figure 1.5 shows how the Lorentz factor grows as the relative speed V of the two frames increases. For speeds that are small compared with the speed of light, $\gamma(V) \approx 1$, and the Lorentz transformations approximate the Galilean transformations provided that x is not too large. As the relative speed of the two frames approaches the speed of light, however, the Lorentz factor grows rapidly and so do the discrepancies between the Galilean and Lorentz transformations.

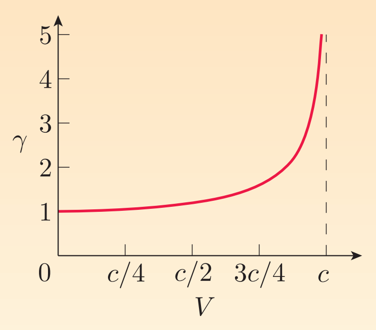

Figure 1.5 Plot of the Lorentz factor, $\gamma(V) = 1/\sqrt{1 - V^2/c^2}$. The factor is close to 1 for speeds much smaller than the speed of light, but increases rapidly as V approaches c. Note that $\gamma > 1$ for all values of V.

Exercise 1.2 Compute the Lorentz factor $\gamma(V)$ when the relative speed V is (a) 10% of the speed of light, and (b) 90% of the speed of light. ■

The Lorentz transformations are so important in special relativity that you will see them written in many different ways. They are often presented in matrix form, as

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma(V) & -\gamma(V)V/c & 0 & 0 \\ -\gamma(V)V/c & \gamma(V) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \tag{1.10}$$

You should convince yourself that this matrix multiplication gives equations equivalent to the Lorentz transformations. (The equation for transforming the time coordinate is multiplied by c.) We can also represent this relationship by the equation

$$[x'^\mu] = [\Lambda^\mu{}_\nu][x^\nu], \tag{1.11}$$

where we use the symbol $[x^\mu]$ to represent the column vector with components $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$, and the symbol $[\Lambda^\mu{}_\nu]$ to represent the **Lorentz transformation matrix**

$$[\Lambda^\mu{}_\nu] \equiv \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix}$$

$$= \begin{pmatrix} \gamma(V) & -\gamma(V)V/c & 0 & 0 \\ -\gamma(V)V/c & \gamma(V) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.12}$$

At this stage, when dealing with an individual matrix element $\Lambda^\mu{}_\nu$, you can simply regard the first index as indicating the row to which it belongs and the second index as indicating the column. It then makes sense that each of the elements x^μ in the column vector $[x^\mu]$ should have a raised index. However, as you will see later, in the context of relativity the positioning of these indices actually has a much greater significance.

The quantity $[x^\mu]$ is sometimes called the **four-position** since its four components (ct, x, y, z) describe the position of the event in time and space. Note that by using ct to convey the time information, rather than just t, all four components of the four-position are measured in units of distance. Also note that the Greek indices μ and ν take the values 0 to 3. It is conventional in special and general relativity to start the indexing of the vectors and matrices from zero, where $x^0 = ct$. This is because the time coordinate has special properties.

Using the individual components of the four-position, another way of writing the Lorentz transformation is in terms of summations:

$$x'^\mu = \sum_{\nu=0}^{3} \Lambda^\mu{}_\nu \, x^\nu \quad (\mu = 0, 1, 2, 3). \tag{1.13}$$

This one line really represents four different equations, one for each value of μ. When an index is used in this way, it is said to be a **free index**, since we are free to give it any value between 0 and 3, and whatever choice we make indicates a different equation. The index ν that appears in the summation is not free, since whatever value we choose for μ, we are required to sum over all possible values of ν to obtain the final equation. This means that we could replace all appearances of ν by some other index, α say, without actually changing anything. An index that is summed over in this way is said to be a **dummy index**.

Familiarity with the summation form of the Lorentz transformations is particularly useful when beginning the discussion of general relativity; you will meet many such sums. Before moving on, you should convince yourself that you can easily switch between the use of separate equations, matrices (including the use of four-positions) and summations when representing Lorentz transformations.

Given the coordinates of an event in frame S, the Lorentz transformations tell us the coordinates of that same event as observed in frame S′. It is equally important that there is some way to transform coordinates of an event in frame S′ back into the coordinates in frame S. The transformations that perform this task are known as the **inverse Lorentz transformations**.

The inverse Lorentz transformations

$$t = \gamma(V)(t' + Vx'/c^2), \tag{1.14}$$

$$x = \gamma(V)(x' + Vt'), \tag{1.15}$$

$$y = y', \tag{1.16}$$

$$z = z'. \tag{1.17}$$

Note that the only difference between the Lorentz transformations and their inverses is that all the primed and unprimed quantities have been interchanged, and the relative speed of the two frames, V, has been replaced by the quantity $-V$. (This changes the transformations but not the value of the Lorentz factor, which depends only on V^2, so can still be written as $\gamma(V)$.) This relationship between the transformations is expected, since frame S′ is moving with speed V in the positive x-direction as measured in frame S, while frame S is moving with speed V in the negative x'-direction as measured in frame S′. You should confirm that performing a Lorentz transformation and its inverse transformation in succession really does lead back to the original coordinates, i.e. $(ct, x, y, z) \to (ct', x', y', z') \to (ct, x, y, z)$.

● An event occurs at coordinates $(ct = 3\,\text{m},\ x = 4\,\text{m},\ y = 0,\ z = 0)$ in frame S according to an observer O. What are the coordinates of the same event in frame S′ according to an observer O′, moving with speed $V = 3c/4$ in the positive x-direction, as measured in S?

❍ First, the Lorentz factor $\gamma(V)$ should be computed:

$$\gamma(3c/4) = 1/\sqrt{1 - 3^2/4^2} = 4/\sqrt{7}.$$

The new coordinates are then given by the Lorentz transformations:

$$ct' = c\gamma(3c/4)(t - 3x/4c) = (4/\sqrt{7})(3\,\mathrm{m} - 3c \times 4\,\mathrm{m}/4c) = 0\,\mathrm{m},$$
$$x' = \gamma(3c/4)(x - 3tc/4) = (4/\sqrt{7})(4\,\mathrm{m} - 3 \times 3\,\mathrm{m}/4) = \sqrt{7}\,\mathrm{m},$$
$$y' = y = 0\,\mathrm{m},$$
$$z' = z = 0\,\mathrm{m}.$$

Exercise 1.3 The matrix equation

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma(V) & -\gamma(V)V/c \\ -\gamma(V)V/c & \gamma(V) \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}$$

can be inverted to determine the coordinates (ct, x) in terms of (ct', x'). Show that inverting the 2×2 matrix leads to the inverse Lorentz transformations in Equations 1.14 and 1.15. ■

1.2.3 A derivation of the Lorentz transformations

This subsection presents a derivation of the Lorentz transformations that relates the coordinates of an event in two inertial frames, S and S′, that are in standard configuration. It mainly ignores the y- and z-coordinates and just considers the transformation of the t- and x-coordinates of an event. A general transformation relating the coordinates (t', x') of an event in frame S′ to the coordinates (t, x) of the same event in frame S may be written as

$$t' = a_0 + a_1 t + a_2 x + a_3 t^2 + a_4 x^2 + \cdots\,, \qquad (1.18)$$
$$x' = b_0 + b_1 x + b_2 t + b_3 x^2 + b_4 t^2 + \cdots\,, \qquad (1.19)$$

where the dots represent additional terms involving higher powers of x or t.

Now, we know from the definition of standard configuration that the event marking the coincidence of the origins of frames S and S′ has the coordinates $(t, x) = (0, 0)$ in S and $(t', x') = (0, 0)$ in S′. It follows from Equations 1.18 and 1.19 that the constants a_0 and b_0 are zero.

The transformations in Equations 1.18 and 1.19 can be further simplified by the requirement that the observers are using inertial frames of reference. Since Newton's first law must hold in all inertial frames of reference, it is necessary that an object not accelerating in one set of coordinates is also not accelerating in the other set of coordinates. If the higher-order terms in x and t were not zero, then an object observed to have no acceleration in S (such as a spaceship with its thrusters off moving on the line $x = vt$, shown in the upper part of Figure 1.6) would be observed to accelerate in terms of x' and t' (i.e. $x' \neq v't'$, as indicated in the lower part of Figure 1.6). Observer O would report no force on the spaceship, while observer O′ would report some unknown force acting on it. In this way, the two observers would register different laws of physics, violating the first postulate of special relativity. The higher-order terms are therefore inconsistent with the required physics and must be removed, leaving only a linear transformation.

So we expect the special relativistic coordinate transformation between two frames in standard configuration to be represented by linear equations of the form

$$t' = a_1 t + a_2 x, \qquad (1.20)$$
$$x' = b_1 x + b_2 t. \qquad (1.21)$$

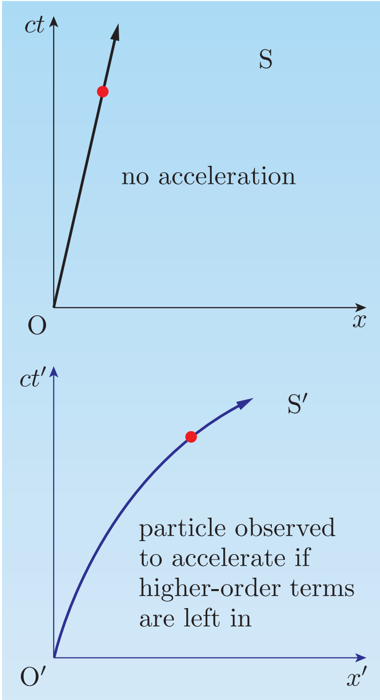

Figure 1.6 Leaving higher-order terms in the coordinate transformations would cause uniform motion in one inertial frame S to be observed as accelerated motion in the other inertial frame S′. These diagrams, in which the vertical axis represents time multiplied by the speed of light, show that if the t^2 terms were left in the transformations, then motion with no acceleration in frame S would be transformed into motion with non-zero acceleration in frame S′. This would imply change in velocity without force in S′, in conflict with Newton's first law.

The remaining task is to determine the coefficients a_1, a_2, b_1 and b_2.

To do this, use is made of known relations between coordinates in both frames of reference. The first step is to use the fact that at any time t, the origin of S′ (which is always at $x' = 0$ in S′) will be at $x = Vt$ in S. It follows from Equation 1.21 that

$$0 = b_1 Vt + b_2 t,$$

from which we see that

$$b_2 = -b_1 V. \tag{1.22}$$

Dividing Equation 1.21 by Equation 1.20, and using Equation 1.22 to replace b_2 by $-b_1 V$, leads to

$$\frac{x'}{t'} = \frac{b_1 x - b_1 Vt}{a_1 t + a_2 x}. \tag{1.23}$$

Now, as a second step we can use the fact that at any time t', the origin of frame S (which is always at $x = 0$ in S) will be at $x' = -Vt'$ in S′. Substituting these values for x and x' into Equation 1.23 gives

$$\frac{-Vt'}{t'} = \frac{-b_1 Vt}{a_1 t}, \tag{1.24}$$

from which it follows that

$$b_1 = a_1.$$

If we now substitute $a_1 = b_1$ into Equation 1.23 and divide the numerator and denominator on the right-hand side by t, then

$$\frac{x'}{t'} = \frac{b_1(x/t) - Vb_1}{b_1 + a_2(x/t)}. \tag{1.25}$$

As a third step, the coefficient a_2 can be found using the principle of the constancy of the speed of light. A pulse of light emitted in the positive x-direction from $(ct = 0,\ x = 0)$ has speed $c = x'/t'$ and also $c = x/t$. Substituting these values into Equation 1.25 gives

$$c = \frac{b_1 c - Vb_1}{b_1 + a_2 c},$$

which can be rearranged to give

$$a_2 = -Vb_1/c^2 = -Va_1/c^2. \tag{1.26}$$

Now that a_2, b_1 and b_2 are known in terms of a_1, the coordinate transformations between the two frames can be written as

$$t' = a_1(t - Vx/c^2), \tag{1.27}$$

$$x' = a_1(x - Vt). \tag{1.28}$$

All that remains for the fourth step is to find an expression for a_1. To do this, we first write down the inverse transformations to Equations 1.27 and 1.28, which are found by exchanging primes and replacing V by $-V$. (We are implicitly assuming that a_1 depends only on some even power of V.) This gives

$$t = a_1(t' + Vx'/c^2), \tag{1.29}$$

$$x = a_1(x' + Vt'). \tag{1.30}$$

Substituting Equations 1.29 and 1.30 into Equation 1.28 gives

$$x' = a_1 \left(a_1(x' + Vt') - Va_1 \left(t' + \frac{V}{c^2} x' \right) \right).$$

The second and third terms involving $a_1 V t'$ cancel in this expression, leaving an expression in which the x' cancels on both sides:

$$x' = a_1^2 \left(1 - \frac{V^2}{c^2} \right) x'.$$

By rearranging this equation and taking the positive square root, the coefficient a_1 is determined to be

$$a_1 = \frac{1}{\sqrt{1 - V^2/c^2}}. \tag{1.31}$$

Thus a_1 is seen to be the Lorentz factor $\gamma(V)$, which completes the derivation.

Some further arguments allow the Lorentz transformations to be extended to one time and three space dimensions. There can be no y and z contributions to the transformations for t' and x' since the y- and z-axes could be oriented in any of the perpendicular directions without affecting the events on the x-axis. Similarly, there can be no contributions to the transformations for y' and z' from any other coordinates, as space would become distorted in a non-symmetric manner.

1.2.4 Intervals and their transformation rules

Knowing how the coordinates of an event transform from one frame to another, it is relatively simple to determine how the coordinate intervals that separate pairs of events transform. As you will see in the next section, the rules for transforming intervals are often very useful.

Intervals

An **interval** between two events, measured along a specified axis in a given frame of reference, is the difference in the corresponding coordinates of the two events.

To develop transformation rules for intervals, consider the Lorentz transformations for the coordinates of two events labelled 1 and 2:

$$\begin{aligned} t_1' &= \gamma(V)(t_1 - Vx_1/c^2), & x_1' &= \gamma(V)(x_1 - Vt_1), \\ y_1' &= y_1, & z_1' &= z_1 \\ t_2' &= \gamma(V)(t_2 - Vx_2/c^2), & x_2' &= \gamma(V)(x_2 - Vt_2), \\ y_2' &= y_2, & z_2' &= z_2. \end{aligned}$$

Subtracting the transformation equation for t_1' from that for t_2', and subtracting the transformation equation for x_1' from that for x_2', and so on, gives the following **transformation rules for intervals**:

$$\Delta t' = \gamma(V)(\Delta t - V\Delta x/c^2), \quad (1.32)$$
$$\Delta x' = \gamma(V)(\Delta x - V\Delta t), \quad (1.33)$$
$$\Delta y' = \Delta y, \quad (1.34)$$
$$\Delta z' = \Delta z, \quad (1.35)$$

where $\Delta t = t_2 - t_1$, $\Delta x = x_2 - x_1$, $\Delta y = y_2 - y_1$ and $\Delta z = z_2 - z_1$ denote the various time and space intervals between the events. The inverse transformations for intervals have the same form, with V replaced by $-V$:

$$\Delta t = \gamma(V)(\Delta t' + V\Delta x'/c^2), \quad (1.36)$$
$$\Delta x = \gamma(V)(\Delta x' + V\Delta t'), \quad (1.37)$$
$$\Delta y = \Delta y', \quad (1.38)$$
$$\Delta z = \Delta z'. \quad (1.39)$$

The transformation rules for intervals are useful because they depend only on coordinate differences and not on the specific locations of events in time or space.

1.3 Consequences of the Lorentz transformations

In this section, some of the extraordinary consequences of the Lorentz transformations will be examined. In particular, we shall consider the findings of different observers regarding the rate at which a clock ticks, the length of a rod and the simultaneity of a pair of events. In each case, the trick for determining how the relevant property transforms between frames of reference is to carefully specify how intuitive concepts such as length or duration should be defined consistently in different frames of reference. This is most easily done by identifying each concept with an appropriate *interval* between two events: 1 and 2. Once this has been achieved, we can determine which intervals are known and then use the interval transformation rules (Equations 1.32–1.35 and 1.36–1.39) to find relationships between them. The rest of this section will give examples of this process.

1.3.1 Time dilation

One of the most celebrated consequences of special relativity is the finding that 'moving clocks run slow'. More precisely, any inertial observer must observe that the clocks used by another inertial observer, in uniform relative motion, will run slow. Since clocks are merely indicators of the passage of time, this is really the assertion that any inertial observer will find that time passes more slowly for any other inertial observer who is in relative motion. Thus, according to special relativity, if you and I are inertial observers, and we are in uniform relative motion, then I can perform measurements that will show that time is passing more slowly for you and, simultaneously, you can perform measurements that will show that time is passing more slowly for me. Both of us will be right because time is a relative quantity, not an absolute one. To show how this effect follows from the Lorentz transformations, it is essential to introduce clear, unambiguous definitions of the time intervals that are to be related.

Rather than deal with ticking clocks, our discussion here will refer to short-lived sub-nuclear particles of the sort routinely studied at CERN and other particle physics laboratories. For the purpose of the discussion, a short-lived particle is considered to be a point-like object that is created at some event, labelled 1, and subsequently decays at some other event, labelled 2. The time interval between these two events, as measured in any particular inertial frame, is the **lifetime** of the particle in that frame. This interval is analogous to the time between successive ticks of a clock.

We shall consider the lifetime of a particular particle as observed by two different inertial observers O and O′. Observer O uses a frame S that is fixed in the laboratory, in which the particle travels with constant speed V in the positive x-direction. We shall call this the **laboratory frame**. Observer O′ uses a frame S′ that moves with the particle. Such a frame is called the **rest frame** of the particle since the particle is always at rest in that frame. (You can think of the observer O′ as riding on the particle if you wish.)

According to observer O′, the birth and decay of the (stationary) particle happen at the same place, so if event 1 occurs at (t'_1, x'), then event 2 occurs at (t'_2, x'), and the lifetime of the particle will be $\Delta t' = t'_2 - t'_1$. In special relativity, the time between two events measured in a frame in which the events happen at the same position is called the **proper time** between the events and is usually denoted by the symbol $\Delta\tau$. So, in this case, we can say that in frame S′ the intervals of time and space that separate the two events are $\Delta t' = \Delta\tau = t'_2 - t'_1$ and $\Delta x' = 0$.

According to observer O in the laboratory frame S, event 1 occurs at (t_1, x_1) and event 2 at (t_2, x_2), and the lifetime of the particle is $\Delta t = t_2 - t_1$, which we shall call ΔT. Thus in frame S the intervals of time and space that separate the two events are $\Delta t = \Delta T = t_2 - t_1$ and $\Delta x = x_2 - x_1$.

These events and intervals are represented in Figure 1.7, and everything we know about them is listed in Table 1.1. Such a table is helpful in establishing which of the interval transformations will be useful.

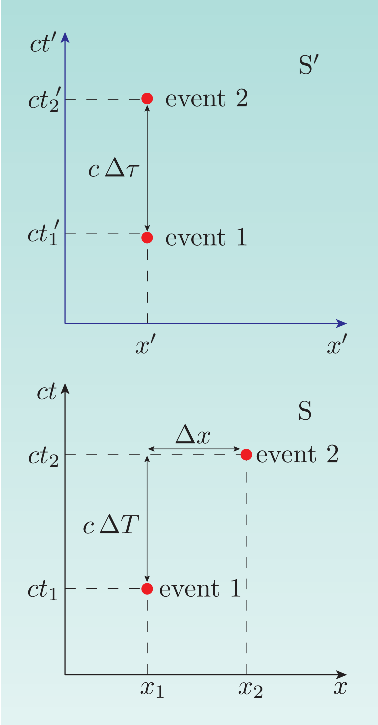

Figure 1.7 Events and intervals for establishing the relation between the lifetime of a particle in its rest frame (S′) and in a laboratory frame (S). Note that we show the coordinate on the vertical axis as 'ct' rather than 't' to ensure that both axes have the dimension of length. To convert time intervals such as $\Delta\tau$ and ΔT to this coordinate, simply multiply them by the constant c.

Table 1.1 A tabular approach to time dilation. The coordinates of the events are listed and the intervals between them worked out, taking account of any known values. The last row is used to show which of the intervals relates to a named quantity (such as the lifetimes ΔT and $\Delta\tau$) or has a known value (such as $\Delta x' = 0$). Any interval that is neither known nor related to a named quantity is shown as a question mark.

Event	S (laboratory)	S′ (rest frame)
2	(t_2, x_2)	(t'_2, x')
1	(t_1, x_1)	(t'_1, x')
Intervals	$(t_2 - t_1, x_2 - x_1)$ $\equiv (\Delta t, \Delta x)$	$(t'_2 - t'_1, 0)$ $\equiv (\Delta t', \Delta x')$
Relation to known intervals	$(\Delta T, ?)$	$(\Delta\tau, 0)$

Four of the interval transformation rules that were introduced in the previous section involve three intervals. But only Equation 1.36 involves the three known intervals. Substituting the known intervals into that equation gives

Figure 1.8 Henri Poincaré (1854–1912).

$\Delta T = \gamma(V)(\Delta\tau + 0)$. Therefore the particle lifetimes measured in S and S$'$ are related by

$$\Delta T = \gamma(V)\,\Delta\tau. \tag{1.40}$$

Since $\gamma(V) > 1$, this result tells us that the particle is observed to live longer in the laboratory frame than it does in its own rest frame. This is an example of the effect known as **time dilation**. A process that occupies a (proper) time $\Delta\tau$ in its own rest frame has a longer duration ΔT when observed from some other frame that moves relative to the rest frame. If the process is the ticking of a clock, then a consequence is that moving clocks will be observed to run slow.

The time dilation effect has been demonstrated experimentally many times. It provides one of the most common pieces of evidence supporting Einstein's theory of special relativity. If it did not exist, many experiments involving short-lived particles, such as *muons*, would be impossible, whereas they are actually quite routine.

It is interesting to note that the French mathematician Henri Poincaré (Figure 1.8) proposed an effect similar to time dilation shortly before Einstein formulated special relativity.

Exercise 1.4 A particular muon lives for $\Delta\tau = 2.2\,\mu$s in its own rest frame. If that muon is travelling with speed $V = 3c/5$ relative to an observer on Earth, what is its lifetime as measured by that observer? ■

1.3.2 Length contraction

There is another curious relativistic effect that relates to the length of an object observed from different frames of reference. For the sake of simplicity, the object that we shall consider is a rod, and we shall start our discussion with a definition of the rod's length that applies whether or not the rod is moving.

In any inertial frame of reference, the **length** of a rod is the distance between its end-points at a single time as measured in that frame.

Thus, in an inertial frame S in which the rod is oriented along the x-axis and moves along that axis with constant speed V, the length L of the rod can be related to two events, 1 and 2, that happen at the ends of the rod at the same time t. If event 1 is at (t, x_1) and event 2 is at (t, x_2), then the length of the rod, as measured in S at time t, is given by $L = \Delta x = x_2 - x_1$.

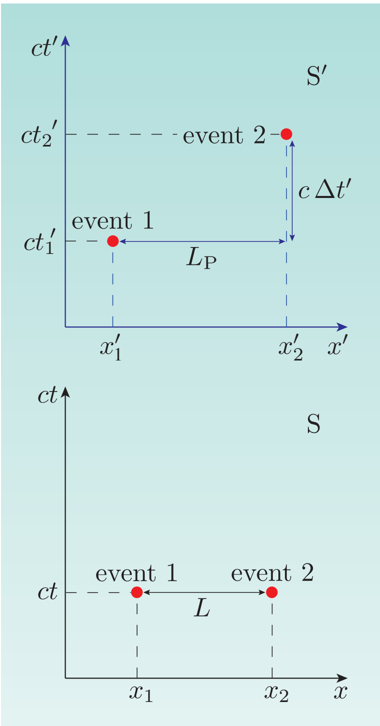

Figure 1.9 Events and intervals for establishing the relation between the length of a rod in its rest frame (S$'$) and in a laboratory frame (S).

Now consider these same two events as observed in an inertial frame S$'$ in which the rod is oriented along the x'-axis but is always at rest. In this case we still know that event 1 and event 2 occur at the end-points of the rod, but we have no reason to suppose that they will occur at the same time, so we shall describe them by the coordinates (t_1', x_1') and (t_2', x_2'). Although these events may not be simultaneous, we know that in frame S$'$ the rod is not moving, so its end-points are always at x_1' and x_2'. Consequently, we can say that the length of the rod in its own rest frame — a quantity sometimes referred to as the **proper length** of the rod and denoted $L_{\rm P}$ — is given by $L_{\rm P} = \Delta x' = x_2' - x_1'$.

These events and intervals are represented in Figure 1.9, and everything we know about them is listed in Table 1.2.

Table 1.2 Events and intervals for length contraction.

Event	S (laboratory)	S′ (rest frame)
2	(t, x_2)	(t'_2, x'_2)
1	(t, x_1)	(t'_1, x'_1)
Intervals	$(0, x_2 - x_1)$ $\equiv (\Delta t, \Delta x)$	$(t'_2 - t'_1, x'_2 - x'_1)$ $\equiv (\Delta t', \Delta x')$
Relation to known intervals	$(0, L)$	$(?, L_{\mathrm{P}})$

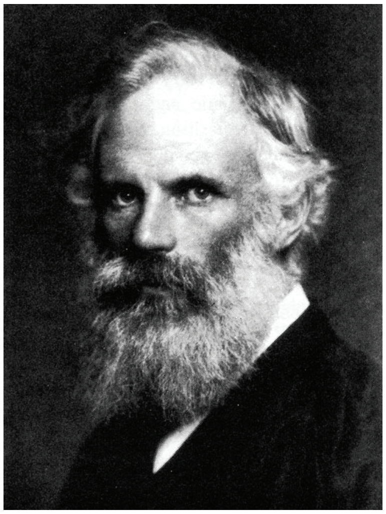

Figure 1.10 George Fitzgerald (1851–1901) was an Irish physicist interested in electromagnetism. He was influential in understanding that length contracts.

On this occasion, the one unknown interval is $\Delta t'$, so the interval transformation rule that relates the three known intervals is Equation 1.33. Substituting the known intervals into that equation gives $L_{\mathrm{P}} = \gamma(V)(L - 0)$. So the lengths measured in S and S′ are related by

$$L = L_{\mathrm{P}}/\gamma(V). \tag{1.41}$$

Since $\gamma(V) > 1$, this result tells us that the rod is observed to be shorter in the laboratory frame than in its own rest frame. In short, moving rods contract. This is an example of the effect known as **length contraction**. The effect is not limited to rods. Any moving body will be observed to contract along its direction of motion, though it is particularly important in this case to remember that this does not mean that it will necessarily be *seen* to contract. There is a substantial body of literature relating to the visual appearance of rapidly moving bodies, which generally involves factors apart from the observed length of the body.

Length contraction is sometimes known as *Lorentz–Fitzgerald contraction* after the physicists (Figure 1.4 and Figure 1.10) who first suggested such a phenomenon, though their interpretation was rather different from that of Einstein.

Exercise 1.5 There is an alternative way of defining length in frame S based on two events, 1 and 2, that happen at *different* times in that frame. Suppose that event 1 occurs at $x = 0$ as the front end of the rod passes that point, and event 2 also occurs at $x = 0$ but at the later time when the rear end passes. Thus event 1 is at $(t_1, 0)$ and event 2 is at $(t_2, 0)$. Since the rod moves with uniform speed V in frame S, we can define the length of the rod, as measured in S, by the relation $L = V(t_2 - t_1)$. Use this alternative definition of length in frame S to establish that the length of a moving rod is less than its proper length. (The events are represented in Figure 1.11.) ■

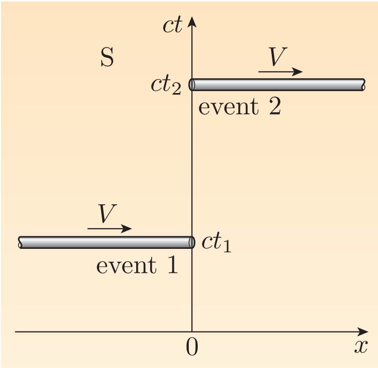

Figure 1.11 An alternative set of events that can be used to determine the length of a uniformly moving rod.

1.3.3 The relativity of simultaneity

It was noted in the discussion of length contraction that two events that occur at the same time in one frame do not necessarily occur at the same time in another frame. Indeed, looking again at Figure 1.9 and Table 1.2 but now calling on the interval transformation rule of Equation 1.32, it is clear that if the events 1 and 2 are observed to occur at the same time in frame S (so $\Delta t = 0$) but are separated by a distance L along the x-axis, then in frame S′ they will be separated by the time

$$\Delta t' = -\gamma(V)VL/c^2.$$

Two events that occur at the same time in some frame are said to be **simultaneous** in that frame. The above result shows that the condition of being simultaneous is a relative one not an absolute one; two events that are simultaneous in one frame are not necessarily simultaneous in every other frame. This consequence of the Lorentz transformations is referred to as the **relativity of simultaneity**.

1.3.4 The Doppler effect

A physical phenomenon that was well known long before the advent of special relativity is the **Doppler effect**. This accounts for the difference between the emitted and received frequencies (or wavelengths) of radiation arising from the relative motion of the emitter and the receiver. You will have heard an example of the Doppler effect if you have listened to the siren of a passing ambulance: the frequency of the siren is higher when the ambulance is approaching (i.e. travelling towards the receiver) than when it is receding (i.e. travelling away from the receiver).

Astronomers routinely use the Doppler effect to determine the speed of approach or recession of distant stars. They do this by measuring the received wavelengths of narrow lines in the star's spectrum, and comparing their results with the proper wavelengths of those lines that are well known from laboratory measurements and represent the wavelengths that would have been emitted in the star's rest frame.

Despite the long history of the Doppler effect, one of the consequences of special relativity was the recognition that the formula that had traditionally been used to describe it was wrong. We shall now obtain the correct formula.

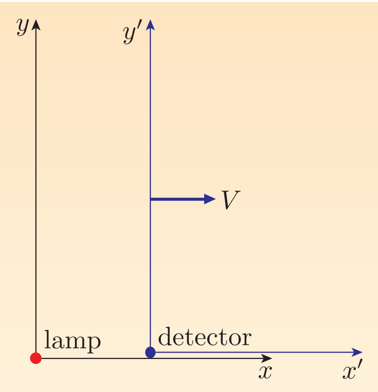

Figure 1.12 The Doppler effect arises from the relative motion of the emitter and receiver of radiation.

Consider a lamp at rest at the origin of an inertial frame S emitting electromagnetic waves of proper frequency f_{em} as measured in S. Now suppose that the lamp is observed from another inertial frame S′ that is in standard configuration with S, moving away at constant speed V (see Figure 1.12). A detector fixed at the origin of S′ will show that the radiation from the receding lamp is received with frequency f_{rec} as measured in S′. Our aim is to find the relationship between f_{rec} and f_{em}.

The emitted waves have regularly positioned *nodes* (points of zero disturbance) that are separated by a proper wavelength $\lambda_{\text{em}} = f_{\text{em}}/c$ as measured in S. In that frame the time interval between the emission of one node and the next, Δt, represents the proper period of the wave, T_{em}, so we can write $\Delta t = T_{\text{em}} = 1/f_{\text{em}}$.

Due to the phenomenon of time dilation, an observer in frame S′ will find that the time separating the emission of successive nodes is $\Delta t' = \gamma(V)\,\Delta t$. However, this is not the time that separates the arrival of those nodes at the detector because the detector is moving away from the emitter at a constant rate. In fact, during the interval $\Delta t'$ the detector will increase its distance from the emitter by $V\Delta t'$ as measured in S′, and this will cause the reception of the two nodes to be separated by a total time $\Delta t' + V\Delta t'/c$ as measured in S′. This represents the received period of the wave and is therefore the reciprocal of the received frequency, so we can write

$$\frac{1}{f_{\text{rec}}} = \Delta t' + \frac{V\Delta t'}{c} = \gamma(V)\,\Delta t\left(1 + \frac{V}{c}\right).$$

We can now identify Δt with the reciprocal of the emitted frequency and use the identity $\gamma(V) = 1/\sqrt{(1 - V/c)(1 + V/c)}$ to write

$$\frac{1}{f_{\text{rec}}} = \frac{1}{f_{\text{em}}} \frac{1}{\sqrt{(1 - V/c)(1 + V/c)}} \left(1 + \frac{V}{c}\right),$$

which can be rearranged to give

$$f_{\text{rec}} = f_{\text{em}} \sqrt{\frac{c - V}{c + V}}. \tag{1.42}$$

This shows that the radiation received from a receding source will have a frequency that is less than the proper frequency with which the radiation was emitted. It follows that the received wavelength $\lambda_{\text{rec}} = c/f_{\text{rec}}$ will be greater than the proper wavelength λ_{em}. Consequently, the spectral lines seen in the light of receding stars will be shifted towards the red end of the spectrum; a phenomenon known as **redshift** (see Figure 1.13). In a similar way, the spectra of approaching stars will be subject to a **blueshift** described by an equation similar to Equation 1.42 but with V replaced by $-V$ throughout. The correct interpretation of these **Doppler shifts** is of great importance.

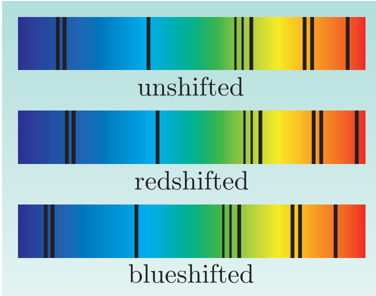

Figure 1.13 Spectral lines are redshifted (that is, reduced in frequency) when the source is receding, and blueshifted (increased in frequency) when the source is approaching.

Exercise 1.6 Some astronomers are studying an unusual phenomenon, close to the centre of our galaxy, involving a jet of material containing sodium. The jet is moving almost directly along the line between the Earth and the galactic centre. In a laboratory, a stationary sample of sodium vapour absorbs light of wavelength $\lambda = 5850 \times 10^{-10}$ m. Spectroscopic studies show that the wavelength of the sodium absorption line in the jet's spectrum is $\lambda' = 4483 \times 10^{-10}$ m. Is the jet approaching or receding? What is the speed of the jet relative to Earth? (Note that the main challenge in this question is the mathematical one of using Equation 1.42 to obtain an expression for V in terms of λ/λ'.) ■

1.3.5 The velocity transformation

Suppose that an object is observed to be moving with velocity $\boldsymbol{v} = (v_x, v_y, v_z)$ in an inertial frame S. What will its velocity be in a frame S′ that is in standard configuration with S, travelling with uniform speed V in the positive x-direction? The Galilean transformation would lead us to expect $\boldsymbol{v}' = (v_x - V, v_y, v_z)$, but we know that is not consistent with the observed behaviour of light. Once again we shall use the interval transformation rules that follow directly from the Lorentz transformations to find the velocity transformation rule according to special relativity.

We know from Equations 1.32 and 1.33 that the time and space intervals between two events 1 and 2 that occur on the x-axis in frame S, transform according to

$$\Delta t' = \gamma(V)(\Delta t - V\Delta x/c^2),$$
$$\Delta x' = \gamma(V)(\Delta x - V\Delta t).$$

Dividing the second of these equations by the first gives

$$\frac{\Delta x'}{\Delta t'} = \frac{\gamma(V)(\Delta x - V\Delta t)}{\gamma(V)(\Delta t - V\Delta x/c^2)}.$$

Dividing the upper and lower expressions on the right-hand side of this equation by Δt, and cancelling the Lorentz factors, gives

$$\frac{\Delta x'}{\Delta t'} = \frac{(\Delta x/\Delta t - V)}{(1 - (\Delta x/\Delta t)V/c^2)}.$$

Now, if we suppose that the two events that we are considering are very close together — indeed, if we consider the limit as Δt and Δx go to zero — then the quantities $\Delta x/\Delta t$ and $\Delta x'/\Delta t'$ will become the instantaneous velocity components v_x and v'_x of a moving object that passes through the events 1 and 2. Extending these arguments to three dimensions by considering events that are not confined to the x-axis leads to the following **velocity transformation** rules:

$$v'_x = \frac{v_x - V}{1 - v_x V/c^2}, \tag{1.43}$$

$$v'_y = \frac{v_y}{\gamma(V)(1 - v_x V/c^2)}, \tag{1.44}$$

$$v'_z = \frac{v_z}{\gamma(V)(1 - v_x V/c^2)}. \tag{1.45}$$

These equations may look rather odd at first sight but they make good sense in the context of special relativity. When v_x and V are small compared to the speed of light c, the term $v_x V/c^2$ is very small and the denominator is approximately 1. In such cases, the Galilean velocity transformation rule, $v'_x = v_x - V$, is recovered as a low-speed approximation to the special relativistic result. At high speeds the situation is even more interesting, as the following question will show.

● An observer has established that two objects are receding in opposite directions. Object 1 has speed c, and object 2 has speed V. Using the velocity transformation, compute the velocity with which object 1 recedes as measured by an observer travelling on object 2.

○ Let the line along which the objects are travelling be the x-axis of the original observer's frame, S. We can then suppose that a frame of reference S$'$ that has its origin on object 2 is in standard configuration with frame S, and apply the velocity transformation to the velocity components of object 1 with $\boldsymbol{v} = (-c, 0, 0)$ (see Figure 1.14). The velocity transformation predicts that as observed in S$'$, the velocity of object 2 is $\boldsymbol{v}' = (v'_x, 0, 0)$, where

$$v'_x = \frac{v_x - V}{1 - v_x V/c^2} = \frac{-c - V}{1 - (-c)V/c^2} = -c.$$

So, as observed from object 2, object 1 is travelling in the $-x'$-direction at the speed of light, c. This was inevitable, since the second postulate of special relativity (which was used in the derivation of the Lorentz transformations) tells us that all observers agree about the speed of light. It is nonetheless pleasing to see how the velocity transformation delivers the required result in this case. It is worth noting that this result does not depend on the value of V.

Exercise 1.7 According to an observer on a spacestation, two spacecraft are moving away, travelling in the same direction at different speeds. The nearer spacecraft is moving at speed $c/2$, the further at speed $3c/4$. What is the speed of one of the spacecraft as observed from the other? ■

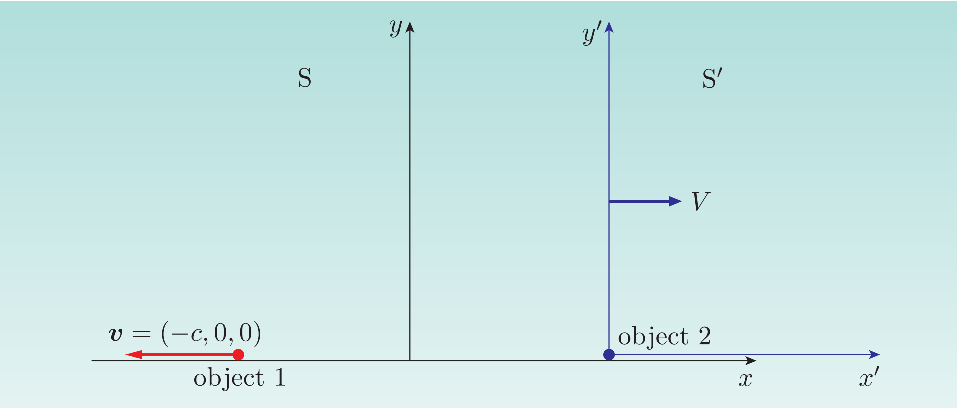

Figure 1.14 Two objects move in opposite directions along the x-axis of frame S. Object 1 travels with speed c; object 2 travels with speed V and is the origin of a second frame of reference S′.

Figure 1.15 Hermann Minkowski (1864–1909) was one of Einstein's mathematics teachers at the Swiss Federal Polytechnic in Zurich. In 1907 he moved to the University of Göttingen, and while there he introduced the idea of spacetime. Einstein was initially unimpressed but later acknowledged his indebtedness to Minkowski for easing the transition from special to general relativity.

1.4 Minkowski spacetime

In 1908 Einstein's former mathematics teacher, Hermann Minkowski (Figure 1.15), gave a lecture in which he introduced the idea of **spacetime**. He said in the lecture: 'Henceforth space by itself, and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality'. This section concerns that four-dimensional union of space and time, the set of all possible events, which is now called **Minkowski spacetime**.

1.4.1 Spacetime diagrams, lightcones and causality

We have already seen how the Lorentz transformations lead to some very counter-intuitive consequences. This subsection introduces a graphical tool known as a **spacetime diagram** or a **Minkowski diagram** that will help you to visualize events in Minkowski spacetime and thereby develop a better intuitive understanding of relativistic effects. The spacetime diagram for a frame of reference S is usually presented as a plot of ct against x, and each point on the diagram represents a possible event as observed in frame S. The y- and z-coordinates are usually ignored.

Given two inertial frames, S and S′, in standard configuration, it is instructive to plot the ct'- and x'-axes of frame S′ on the spacetime diagram for frame S. The x'-axis of frame S′ is defined by the set of events for which $ct' = 0$, and the ct'-axis is defined by the set of events for which $x' = 0$. The coordinates of these events in S are related to their coordinates in S′ by the following Lorentz transformations. (Note that the time transformation of Equation 1.5 has been multiplied by c so that each coordinate can be measured in units of length.)

$$ct' = \gamma(V)(ct - Vx/c),$$
$$x' = \gamma(V)(x - Vt).$$

Setting $ct' = 0$ in the first of these equations gives $0 = \gamma(V)(ct - Vx/c)$. This shows that in the spacetime diagram for frame S, the ct'-axis of frame S′ is represented by the line $ct = (V/c)x$, a straight line through the origin with gradient V/c. Similarly, setting $x' = 0$ in the second transformation equation gives $0 = \gamma(V)(x - Vt)$, showing that the x'-axis of frame S$'$ is represented by the line $ct = (c/V)x$, a straight line through the origin with gradient c/V in the spacetime diagram of S. These lines are shown in Figure 1.16.

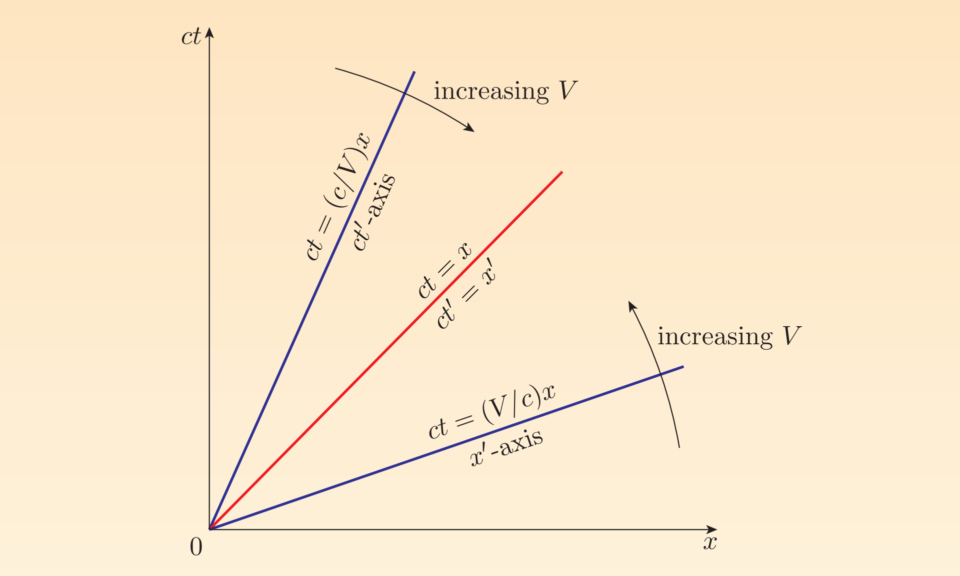

Figure 1.16 The spacetime diagram of frame S, showing the events that make up the ct'- and x'-axes of frame S$'$, and the path of a light ray that passes through the origin.

There is another feature of interest in the diagram. The straight line through the origin of gradient 1 links all the events where $x = ct$ and thus shows the path of a light ray that passes through $x = 0$ at time $t = 0$. Using the inverse Lorentz transformations shows that this line also passes through all the events where $\gamma(V)(x' + Vt') = \gamma(V)(ct' + Vx'/c)$, that is (after some cancelling and rearranging), where $x' = ct'$. So the line of gradient 1 passing through the origin also represents the path of a light ray that passes through the origin of frame S$'$ at $t' = 0$. In fact, any line with gradient 1 on a spacetime diagram must always represent the possible path of a light ray, and thanks to the second postulate of special relativity, we can be sure that all observers will agree about that.

As the relative speed V of the frames S and S$'$ increases, the lines representing the x'- and ct'-axes of S$'$ close in on the line of gradient 1 from either side, rather like the closing of a clapper board. This behaviour reflects the fact that Lorentz transformations will generally alter the coordinates of events but will not change the behaviour of light on which all observers must agree.

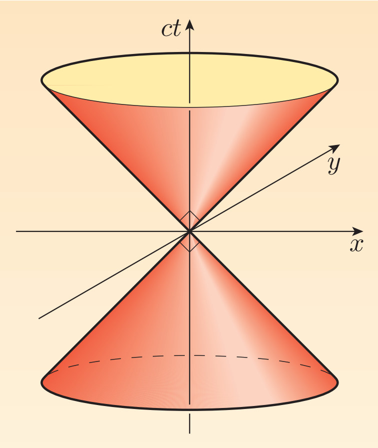

Figure 1.17 In three dimensions (one time and two space) it becomes clear that a line of gradient 1 in a spacetime diagram is part of a lightcone.

In the somewhat unusual case when we include a second spatial axis (the y-axis, say) in the spacetime diagram, the original line of gradient 1 is seen to be part of a cone, as indicated in Figure 1.17. This cone, which connects the event at the origin to all those events, past and future, that might be linked to it by a signal travelling at the speed of light, is an example of a **lightcone**. A horizontal slice (at ct = constant) through the (pseudo) three-dimensional diagram at any particular time shows a circle, but in a fully four-dimensional diagram with all three spatial axes included, such a fixed-time slice would be a sphere, and would represent a spherical shell of light surrounding the origin. At times earlier than $t = 0$, the shell would represent incoming light signals closing in on the origin. At times later than $t = 0$, the shell would represent outgoing light signals travelling away from the origin. Although observers O and O′, using frames S and S′, would not generally agree about the coordinates of events, they would agree about which events were on the lightcone, which were inside the lightcone and which were outside. This agreement between observers makes lightcones very useful in discussions about which events might cause, or be caused by, other events.

Going back to an ordinary two-dimensional spacetime diagram of the kind shown in Figure 1.18, it is straightforward to read off the coordinates of an event in frame S or in frame S′. The event 1 in the diagram clearly has coordinates (ct_1, x_1) in frame S. In frame S′, it has a different set of coordinates. These can be determined by drawing construction lines *parallel* to the lines representing the primed axes. Where a construction line parallel to one primed axis intersects the other primed axis, the coordinate can be found. By doing this on both axes, both coordinates are found. In the case of Figure 1.18, the dashed construction lines show that, as observed in frame S′, event 1 occurs at the same time as event 2, and at the same position as event 3.

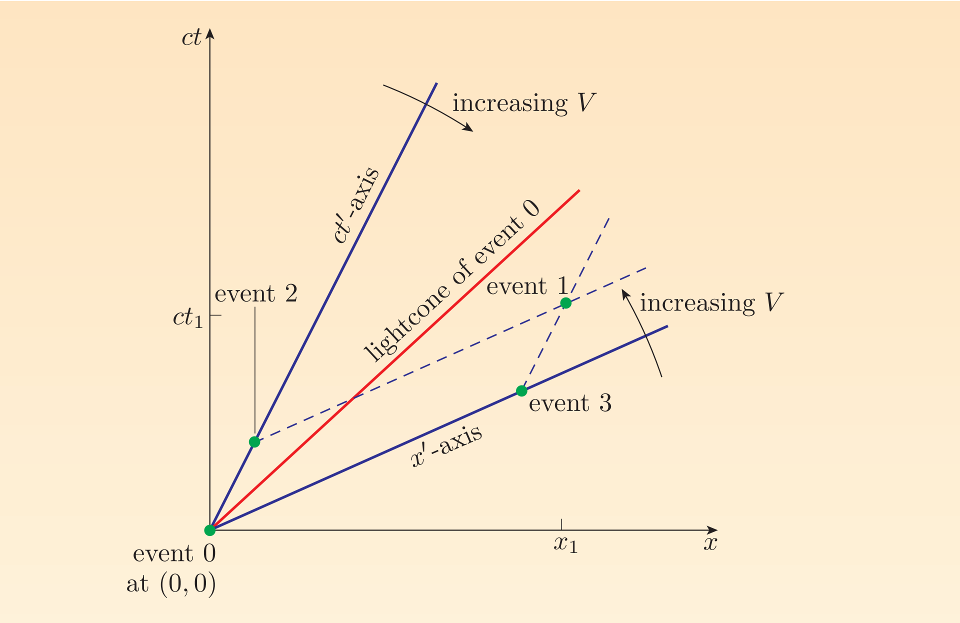

Figure 1.18 A spacetime diagram for frame S with four events, 0, 1, 2 and 3. Event coordinates in S′ can be found by drawing construction lines parallel to the appropriate axes.

Another lesson that can be drawn from Figure 1.18 concerns the order of events. Starting from the bottom of the ct-axis and working upwards, it is clear that in frame S, the four events occur in the order 0, 2, 3 and 1. But it is equally clear from the dashed construction lines that in frame S′, event 3 happens at the same time as event 0 (they are simultaneous in S′), and both happen at an earlier time than event 2 and event 1, which are also simultaneous in S′. This illustrates the relativity of simultaneity, but more importantly it also shows that the order of events 2 and 3 will be different for observers O and O′.

At first sight it is quite shocking to learn that the relative motion of two observers can reverse the order in which they observe events to happen. This has the potential to overthrow our normal notion of **causality**, the principle that all observers must agree that any effect is *preceded* by its cause. It is easy to imagine observing the pressing of a plunger and then observing the explosion that it causes. It would be very shocking if some other observer, simply by moving sufficiently fast in the right direction, was able to observe the explosion first and then the pressing of the plunger that caused it. (It is important to remember that we are discussing observing, not seeing.)

Fortunately, such an overthrow of causality is not permitted by special relativity, *provided that we do not allow signals to travel at speeds greater than* c. Although observers will disagree about the order of some events, they will not disagree about the order of any two events that might be linked by a light signal or any signal that travels at less than the speed of light. Such events are said to be **causally related**.

To see how the order of causally related events is preserved, look again at Figure 1.18, noting that all the events that are causally related to event 0 are contained within its lightcone, and that includes event 2. Events that are not causally related to event 0, such as event 1 and event 3, are outside the lightcone of event 0 and could only be linked to that event by signals that travel faster than light. Now, remember that as the relative speed V of the observers O and O′ increases, the line representing the ct'-axis closes in on the lightcone. As a result, there will not be any value of V that allows the causally related events 0 and 2 to change their order. Event 2 will always be at a higher value of ct' than event 0. However, when you examine the corresponding behaviour of events 0 and 3, which are not causally related, the conclusion is quite different. Figure 1.18 shows the condition in which event 0 and event 3 occur at the same time $t' = 0$, according to O′. When O and O′ have a lower relative speed, event 3 occurs after event 0, but as V increases and the line representing the x'-axis (where all events occur at $ct' = 0$) closes in on the lightcone, we see that there will be a value of V above which the order of event 0 and event 3 is reversed.

So, if event 0 represents the pressing of a plunger and event 2 and event 3 represent explosions, all observers will agree that event 0 might have caused event 2, which happened later. However, those same observers will not agree about the order of event 0 and event 3, though they will agree that event 0 could not have *caused* event 3 unless bodies or signals can travel faster than light. It is the desire to preserve causal relationships that is the basis for the requirement that no material body or signal of any kind should be able to travel faster than light.

- Is event 1 in Figure 1.18 causally related to event 0? Is event 1 causally related to event 3? Justify your answers.

- Event 1 is outside the lightcone of event 0, so the two cannot be causally related. The diagram does not show the lightcone of event 3, but if you imagine a line of gradient 1, parallel to the shown lightcone, passing through event 3, it is clear that event 1 is inside the lightcone of event 3, so those two events are causally related. The earlier event may have caused the later one, and all observers will agree about that.

An important lesson to learn from this question is the significance of drawing lightcones for events other than those at the origin. Every event has a lightcone, and that lightcone is of great value in determining causal relationships.

1.4.2 Spacetime separation and the Minkowski metric

In three-dimensional space, the separation between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) can be conveniently described by the square of the distance Δl between them:

$$(\Delta l)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2, \tag{1.46}$$

where $\Delta x = x_2 - x_1$, $\Delta y = y_2 - y_1$ and $\Delta z = z_2 - z_1$. This quantity has the useful property of being unchanged by rotations of the coordinate system. So, if we choose to describe the points using a new coordinate system with axes x', y' and z', obtained by rotating the old system about one or more of its axes, then the spatial separation of the two points would still be described by an expression of the form

$$(\Delta l')^2 = (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2, \tag{1.47}$$

and we would find in addition that

$$(\Delta l)^2 = (\Delta l')^2. \tag{1.48}$$

We describe this situation by saying that the spatial separation of two points is **invariant** under rotations of the coordinate system used to describe the positions of the two points.

These ideas can be extended to four-dimensional Minkowski spacetime, where the most useful expression for the **spacetime separation** of two events is the following.

Spacetime separation

$$(\Delta s)^2 = (c\,\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2. \tag{1.49}$$

The reason why this particular form is chosen is that it turns out to be invariant under Lorenz transformations. So, if O and O′ are inertial observers using frames S and S′, they will generally not agree about the coordinates that describe two events 1 and 2, or about the distance or the time that separates them, but they will agree that the two events have an invariant spacetime separation

$$(\Delta s)^2 = (c\,\Delta t)^2 - (\Delta l)^2 = (c\,\Delta t')^2 - (\Delta l')^2 = (\Delta s')^2. \tag{1.50}$$

Exercise 1.8 Two events occur at $(ct_1, x_1, y_1, z_1) = (3, 7, 0, 0)$ m and $(ct_2, x_2, y_2, z_2) = (5, 5, 0, 0)$ m. What is their spacetime separation?

Exercise 1.9 In the case that $\Delta y = 0$ and $\Delta z = 0$, use the interval transformation rules to show that the spacetime separation given by Equation 1.49 really is invariant under Lorentz transformations. ■

A convenient way of writing the spacetime separation is as a summation:

$$(\Delta s)^2 = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu}\,\Delta x^\mu\,\Delta x^\nu, \tag{1.51}$$

where the four quantities Δx^0, Δx^1, Δx^2 and Δx^3 are the components of $[\Delta x^\mu] = (c\,\Delta t, \Delta x, \Delta y, \Delta z)$, and the new quantities $\eta_{\mu\nu}$ that have been introduced are the sixteen components of an entity called the **Minkowski metric**, which can be represented as

$$[\eta_{\mu\nu}] \equiv \begin{pmatrix} \eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} \\ \eta_{10} & \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{20} & \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{30} & \eta_{31} & \eta_{32} & \eta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{1.52}$$

It's worth noting that the Minkowski metric has been shown as a matrix only for convenience; Equation 1.51 is not a matrix equation, though it is a well-defined sum. The important point is that the quantity $[\eta_{\mu\nu}]$ has sixteen components, and from Equation 1.52 you can uniquely identify each of them. The metric provides a valuable reminder of how the spacetime separation is related to the coordinate intervals. Metrics will have a crucial role to play in the rest of this book. The Minkowski metric is just the first of many that you will meet.

The spacetime separation of two events is an important quantity for several reasons. Its sign alone tells us about the possible causal relationship between the events. In fact, we can identify three classes of relationship, corresponding to the cases $(\Delta s)^2 > 0$, $(\Delta s)^2 = 0$ and $(\Delta s)^2 < 0$.

Time-like, light-like and space-like separations

Events with a positive spacetime separation, $(\Delta s)^2 > 0$, are said to be **time-like** separated. Such events are causally related, and there will exist a frame in which the two events happen at the same place but at different times.

Events with a zero (or null) spacetime separation, $(\Delta s)^2 = 0$, are said to be **light-like** separated. Such events are causally related, and all observers will agree that they could be linked by a light signal.

Events with a negative spacetime separation, $(\Delta s)^2 < 0$, are said to be **space-like** separated. Such events are not causally related, and there will exist a frame in which the two events happen at the same time but at different places.

These different kinds of spacetime separation correspond to different regions of spacetime defined by the lightcone of an event. Figure 1.19 shows the lightcone of event 0. All the events that have a time-like separation from event 0 are within the future or past lightcone of event 0; all the events that are light-like separated from event 0 are on its lightcone; and all the events that are space-like separated from event 0 are outside its lightcone. This emphasizes the role that lightcones play in revealing the causal structure of Minkowski spacetime.

Another reason why spacetime separation is important relates to *proper time*. You will recall that in the earlier discussion of time dilation, it was said that the proper time between two events was the time separating those events as measured in a frame where the events happen at the same position. In such a frame, the spacetime separation of the events is $(\Delta s)^2 = c^2(\Delta t)^2 = c^2(\Delta\tau)^2$. However,

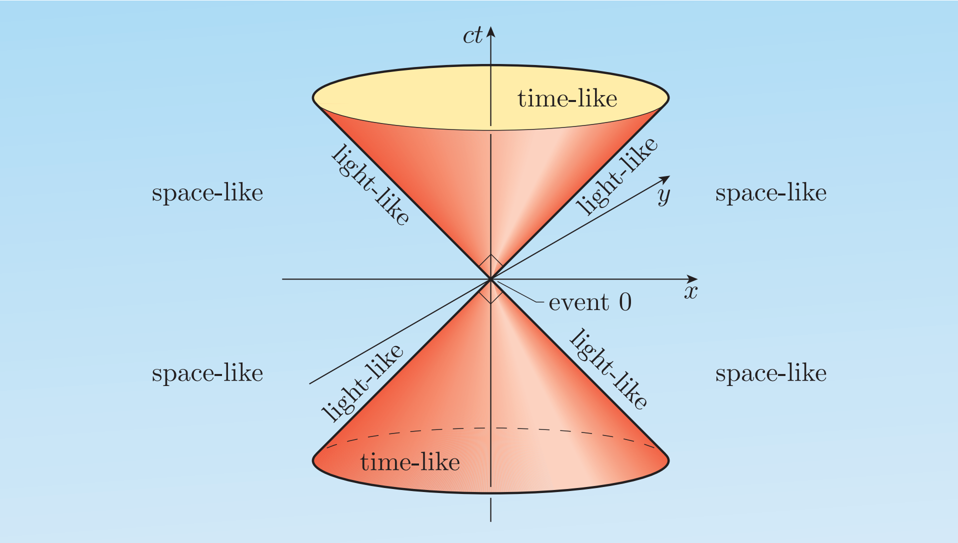

Figure 1.19 Events that are time-like separated from event 0 are found inside its lightcone. Events that are light-like separated are found on the lightcone, and events that are space-like separated from event 0 are outside the lightcone.

since the spacetime separation of events is an invariant quantity, we can use it to determine the proper time between two time-like separated events, irrespective of the frame in which the events are described. For two time-like separated events with positive spacetime separation $(\Delta s)^2$, the proper time $\Delta\tau$ between those two events is given by the following.

Proper time related to spacetime separation

$$(\Delta\tau)^2 = (\Delta s)^2/c^2. \tag{1.53}$$

The relation between proper time and the invariant spacetime separation is extremely useful in special relativity. The reason for this relates to the length of a particle's pathway through four-dimensional Minkowski spacetime. Such a pathway, with all its twists and turns, records the whole history of the particle and is sometimes called its **world-line**. (One well-known relativist called his autobiography *My worldline*.) By adding together the spacetime separations between successive events along a particle's world-line, and dividing the sum by c^2, we can determine the total time that has passed according to a clock carried by the particle. This simple principle will be used to help to explain a troublesome relativistic effect in the next subsection.

In this book, a positive sign will always be associated with the square of the time interval in the spacetime separation, and a negative sign with the spatial intervals. This choice of sign is just a convention, and the opposite set of signs could have been used. The convention used here ensures that the spacetime separation of events on the world-line of an object moving slower than light is positive. Nonetheless, you will find that many authors adopt the opposite convention, so when consulting other works, always pay attention to the sign convention that they are using.

Exercise 1.10 Given two time-like separated events, show that the proper time between those events is the least amount of time that any inertial observer will measure between them. ■

1.4.3 The twin effect

We end this chapter with a discussion of a well-known relativistic effect, the **twin effect**. This caused a great deal of controversy early in the theory's history. It is usually presented as a thought experiment concerning the phenomenon of time dilation. The thought experiment involves two twins, Astra and Terra. The twins are identical in every way, except that Astra likes to travel around very fast in her spaceship, while Terra prefers to stay at home on Earth.

As was demonstrated earlier in this chapter, fast-moving objects are subject to observable time dilation effects. This indicates that if Astra jets off in some fixed direction at close to the speed of light, then, as measured by Terra, she will age more slowly because 'moving clocks run slow'. This is fine — it is just what relativity theory predicts, and agrees with the observed behaviour of high-speed particles. But now suppose that Astra somehow manages to turn around and return to Earth at equally high speed. It seems clear that Terra will again observe that Astra's clock will run slow and will therefore not be surprised to find that on her return, Astra has aged less than her stay-at-home twin Terra.

The supposed problem arises when this process is examined from Astra's point of view. Would it not be the case, some argued, that Astra would observe the same events apart from a reversal of velocities, so that Terra would be the travelling twin and it would be Terra's clock that would be running slow during both parts of the journey? Consequently, shouldn't Astra expect Terra to be the younger when they were reunited? Clearly, it's not possible for each twin to be younger than the other when they meet at the same place, so if the arguments are equally sound, it was said, there must be something wrong with special relativity.

In fact, the arguments are not equally sound. The basic problem is that the presumed symmetry between Terra's view and Astra's view is illusory. It is Astra who would be the younger at the reunion, as will now be explained with the aid of a spacetime diagram and a proper use of spacetime separations in Minkowski space.

The first point to make clear is that although velocity is a purely relative quantity, acceleration is not. According to the first postulate of special relativity, the laws of physics do not distinguish one inertial frame from another, so a traveller in a closed box cannot determine his or her speed by performing a physics experiment. However, such a traveller would certainly be able to feel the effect of any acceleration, as we all know from everyday experience. In order to leave the Solar System, jet around the galaxy and return, Astra must have undergone a change in velocity, and that would involve a detectable acceleration. To a first approximation, Terra does not accelerate (her velocity changes due to the rotation and revolution of the Earth are very small compared with Astra's accelerations). A single inertial frame of reference is sufficient to represent Terra's view of events, but no single *inertial* frame can adequately represent Astra's view. There is no symmetry between these two observers; only Terra is an (approximately) inertial observer.

In order to be clear about what's going on and to avoid the use of non-inertial frames, it is convenient to use three inertial frames when discussing the twin effect. The first is Terra's frame, which we can treat as fixed on a non-rotating, non-revolving Earth. The second, which we shall call Astra's frame, moves at a high but constant speed V relative to Terra's frame. You can think of this as the frame of Astra's spaceship, and you can think of Astra as simply jumping aboard her passing ship at the departure, event 0, when she leaves Terra to begin the outward leg of her journey. The third inertial frame, called Stella's frame, belongs to another space traveller who happens to be approaching Earth at speed V along the same line that Astra leaves along. At some point, Stella's ship will pass Astra's, and at that point we can imagine that Astra jumps from her ship to Stella's ship to make the return leg of her journey. Of course, this is unrealistic since the 'jump' would kill Astra, so you may prefer to imagine that Astra is actually a conscious robot or even that she can somehow 'teleport' from one ship to another. In any case, the important point is that the transfer is abrupt and has no effect on Astra's age.

The event at which Astra makes the transfer to Stella's ship we shall call event 1, and the event at which Astra and Terra are eventually reunited we shall call event 2. Astra's quick transfer from one ship to the other allows us to discuss the essential features of the twin effect without getting bogged down in details about the nature of the acceleration that Astra experiences. It is vital that Astra is accelerated, but exactly how that happens is unimportant. Note that we may treat each of these frames as being in standard configuration with either of the others. We can set up the frames in such a way that the origins of Terra's frame and Astra's frame coincide at event 0, the origins of Astra's frame and Stella's frame coincide at event 1, and the origins of Stella's frame and Terra's frame coincide at event 2.

Figure 1.20 is a spacetime diagram for Terra's frame, showing all these events and making clear the coordinates that Terra assigns to them.

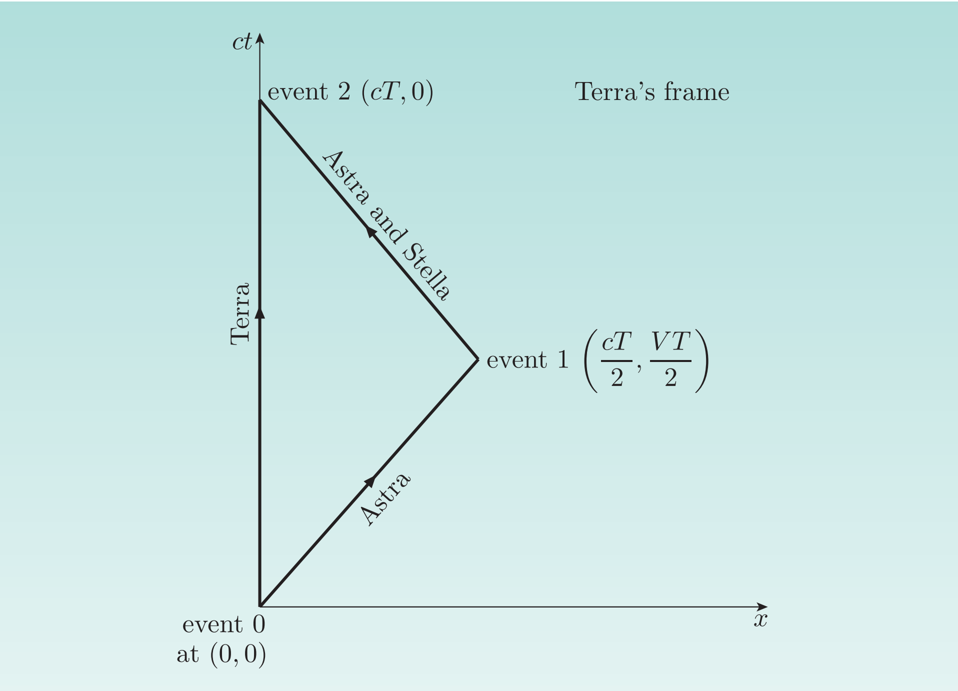

Figure 1.20 A spacetime diagram for Terra's frame, showing the departure, transfer and reunion events together with their coordinates. The t-coordinate has been multiplied by c, as usual.

It is clear from the figure that the proper time between departure and reunion (both of which happen at Terra's location) is T. A little calculation using the relation $(\Delta\tau)^2 = (\Delta s)^2/c^2$ makes it equally clear that the proper time between event 0 and event 1 is given by

$$(\Delta\tau_{0,1})^2 = \frac{(\Delta s_{0,1})^2}{c^2} = \frac{1}{c^2}\left[\left(\frac{cT}{2}\right)^2 - \left(\frac{VT}{2}\right)^2\right]$$
$$= \frac{T^2}{4}\left(1 - \frac{V^2}{c^2}\right) = \left(\frac{T}{2\gamma}\right)^2. \quad (1.54)$$

So

$$\Delta\tau_{0,1} = \frac{T}{2\gamma}. \quad (1.55)$$

Although we have arrived at this result using the coordinates assigned by Terra, it is important to note that proper time is an invariant, so all inertial observers will agree on the proper time between two events no matter how it is calculated.

A similar calculation for the proper time separating event 1 and event 2 shows that

$$\Delta\tau_{1,2} = \frac{T}{2\gamma}. \quad (1.56)$$

So the total proper time that elapses along the world-line followed by Astra is $\Delta\tau_{0,1} + \Delta\tau_{1,2} = T/\gamma$. As expected, this shows that Astra will be the younger twin at the time of the reunion.

How is it possible for Terra and Astra to disagree about the time between events 0 and 2? The answer to this question is that when the whole trip is considered, Astra is *not* an inertial observer; she undergoes an acceleration that Terra does not. Given two time-like separated events, the time that elapses between those events, as measured by an observer present at both events, will depend on the observer's world-line. The total time between the events, measured along the world-line of a non-inertial observer, is generally *less* than the proper time between these events as measured along the world-line of an inertial observer.

The analysis that we have just completed is really sufficient to settle any questions about the twin effect. However, it is still instructive to examine the same events from Astra's frame (which she leaves at event 1). The spacetime diagram for Astra's frame is shown in Figure 1.21. The coordinates of the events have been worked out from those given in Terra's frame using the Lorentz transformations.

- Confirm the coordinate assignments shown in Figure 1.21.

- In Terra's frame, event 0 is at $(ct, x) = (0, 0)$, event 1 at $(cT/2, VT/2)$, and event 2 at $(cT, 0)$. Treating Terra's frame as frame S and Astra's frame as S′, and using the Lorentz transformations $t' = \gamma(t - Vx/c^2)$ and $x' = \gamma(x - Vt)$, it follows immediately that in Astra's frame, event 0 is at $(ct', x') = (0, 0)$, event 1 is at $(ct', x') = (cT/2\gamma, 0)$ (remember that $\gamma(V) = 1/\sqrt{1 - V^2/c^2}$), and event 2 is at $(ct', x') = (c\gamma T, -\gamma VT)$.

Note that again there is a kink in Astra's world-line due to the acceleration that she undergoes. There is no such kink in Terra's world-line since she is an inertial observer. Once again we can work out the proper time that Astra experiences while passing between the three events: this represents the time that would have elapsed according to a clock that Astra carries between each of the events. The proper time between event 0 and event 1 is simply $\Delta\tau_{0,1} = T/2\gamma$, since those events happen at the same place in Astra's frame. The proper time between event 1 and event 2 is given by

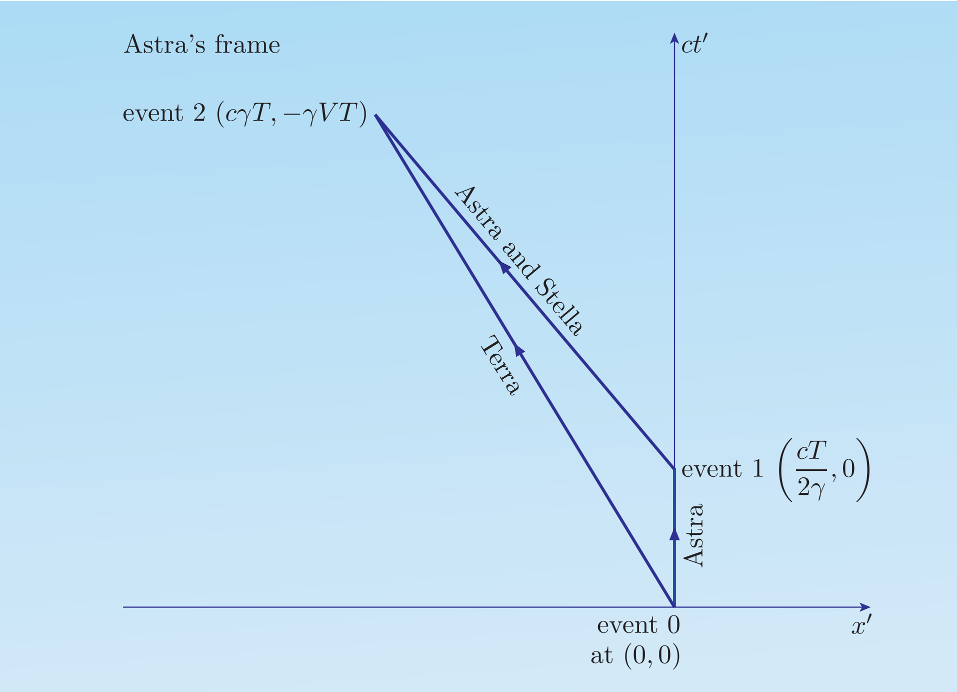

Figure 1.21 A spacetime diagram for Astra's frame, showing the departure, transfer and reunion events with their coordinates. Note that Astra leaves this frame at event 1.

$$\begin{aligned}
(\Delta\tau_{1,2})^2 = \frac{(\Delta s_{1,2})^2}{c^2} &= \frac{1}{c^2}\left[\left(c\gamma T - \frac{cT}{2\gamma}\right)^2 - (-\gamma VT)^2\right] \\
&= T^2\left[\left(\gamma - \frac{1}{2\gamma}\right)^2 - \left(\gamma\frac{V}{c}\right)^2\right] \\
&= T^2\left[\gamma^2 - 1 + \frac{1}{4\gamma^2} - \frac{\gamma^2V^2}{c^2}\right] \\
&= T^2\left[\gamma^2\left(1 - \frac{V^2}{c^2}\right) - 1 + \frac{1}{4\gamma^2}\right].
\end{aligned}$$

Since $\gamma^2(1 - V^2/c^2) = 1$, the above expression simplifies to give

$$\Delta\tau_{1,2} = \frac{T}{2\gamma}.$$

So once again the theory predicts that the time for the round trip recorded by Astra is $\Delta\tau_{0,1} + \Delta\tau_{1,2} = T/\gamma$.

There is one other point to notice using Astra's frame. Time dilation tells us that, as measured in Astra's frame, Terra's clock will be running slow. From Astra's frame, a 1-second tick of Terra's clock will be observed to last γ seconds. But in Astra's frame, it is also the case that the time of the reunion is γT, which is greater than the time of the reunion as observed in Terra's frame. According to an observer who uses Astra's frame, this longer journey time compensates for the slower ticking of Terra's clock, with the result that such an observer will fully expect Terra to have aged by T while Astra herself has aged by only T/γ. Using the coordinates of event 0 and event 2 in Astra's frame, it is easy to confirm that the proper time between them is T, which is another way of stating the same result.

Exercise 1.11 Using the velocity transformation, show that Astra observes the speed of approach of Stella's spaceship to be $2V/(1 + V^2/c^2)$.

Exercise 1.12 Suppose that Terra sends regular time signals towards Astra and Stella at one-second intervals. Write down expressions for the frequency at which Astra receives the signals on the outward and return legs of her journey. ■

Summary of Chapter 1

1. Basic terms in the vocabulary of relativity include: event, frame of reference, inertial frame and observer.
2. A theory of relativity concerns the relationships between observations made by observers in a specified state of relative motion. Special relativity is essentially restricted to inertial observers in uniform relative motion.
3. Einstein based special relativity on two postulates: the principle of relativity (that the laws of physics can be written in the same form in all inertial frames) and the principle of the constancy of the speed of light (that all inertial observers agree that light travels through empty space with the same fixed speed, c, in all directions).
4. Given two inertial frames S and S′ in standard configuration, the coordinates of an event observed in frame S are related to the coordinates of the same event observed in frame S′ by the Lorentz transformations

$$t' = \gamma(V)(t - Vx/c^2), \quad \text{(Eqn 1.5)}$$
$$x' = \gamma(V)(x - Vt), \quad \text{(Eqn 1.6)}$$
$$y' = y, \quad \text{(Eqn 1.7)}$$
$$z' = z, \quad \text{(Eqn 1.8)}$$

where the Lorentz factor is

$$\gamma(V) = \frac{1}{\sqrt{1 - V^2/c^2}}. \quad \text{(Eqn 1.9)}$$

These transformations may also be represented by matrices,

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma(V) & -\gamma(V)V/c & 0 & 0 \\ -\gamma(V)V/c & \gamma(V) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}, \quad \text{(Eqn 1.10)}$$

or as a set of summations

$$x'^{\mu} = \sum_{\nu=0}^{3} \Lambda^{\mu}{}_{\nu}\, x^{\nu} \quad (\mu = 0, 1, 2, 3). \quad \text{(Eqn 1.13)}$$

5. The inverse Lorentz transformations may be written as

$$t = \gamma(V)(t' + Vx'/c^2), \quad \text{(Eqn 1.14)}$$
$$x = \gamma(V)(x' + Vt'), \quad \text{(Eqn 1.15)}$$
$$y = y', \quad \text{(Eqn 1.16)}$$
$$z = z'. \quad \text{(Eqn 1.17)}$$

6. Similar equations describe the transformation of intervals, Δt, Δx, etc., between the two frames.
7. The consequences of special relativity, deduced by considering the transformation of events and intervals, include the following.

   (a) Time dilation:

   $$\Delta T = \gamma(V)\,\Delta\tau. \quad \text{(Eqn 1.40)}$$

   (b) Length contraction:

   $$L = L_{\mathrm{P}}/\gamma(V). \quad \text{(Eqn 1.41)}$$

   (c) The relativity of simultaneity.

   (d) The relativistic Doppler effect (Eqn 1.42):

   $$f_{\mathrm{rec}} = f_{\mathrm{em}}\sqrt{(c+V)/(c-V)} \quad \text{(for an approaching source)},$$
   $$f_{\mathrm{rec}} = f_{\mathrm{em}}\sqrt{(c-V)/(c+V)} \quad \text{(for a receding source)}.$$

   (e) The velocity transformation:

   $$v'_x = \frac{v_x - V}{1 - v_x V/c^2}, \quad \text{(Eqn 1.43)}$$
   $$v'_y = \frac{v_y}{\gamma(V)(1 - v_x V/c^2)}, \quad \text{(Eqn 1.44)}$$
   $$v'_z = \frac{v_z}{\gamma(V)(1 - v_x V/c^2)}. \quad \text{(Eqn 1.45)}$$

8. Four-dimensional Minkowski spacetime contains all possible events.
9. Spacetime diagrams showing events as observed by a particular observer are a valuable tool that can provide pictorial insights into relativistic effects and the structure of Minkowski spacetime.
10. Lightcones are particularly useful for understanding causal relationships between events in Minkowski spacetime.
11. The invariant spacetime separation between two events has the form

$$(\Delta s)^2 = (c\,\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2, \quad \text{(Eqn 1.49)}$$

and may be positive (time-like), zero (light-like) or negative (space-like).

12. The spacetime separation may be conveniently written as

$$(\Delta s)^2 = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu}\,\Delta x^\mu\,\Delta x^\nu, \quad \text{(Eqn 1.51)}$$

where the $\eta_{\mu\nu}$ are the components of the Minkowski metric

$$[\eta_{\mu\nu}] \equiv \begin{pmatrix} \eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} \\ \eta_{10} & \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{20} & \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{30} & \eta_{31} & \eta_{32} & \eta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad \text{(Eqn 1.52)}$$

13. The proper time $\Delta\tau$ between two time-like separated events is given by

$$(\Delta\tau)^2 = (\Delta s)^2/c^2. \quad \text{(Eqn 1.53)}$$

    This is the time that would be recorded on a clock that moves uniformly between the two events.

14. The proper time between two events is an invariant under Lorentz transformations.

15. The time between two time-like separated events, as measured by an observer present at both events, depends on the world-line of the observer. The time between the events, measured by a non-inertial observer, is generally *less* than the proper time between these events as measured by an inertial observer. This is the basis of the twin effect.

Acknowledgements

Grateful acknowledgement is made to the following sources:

Figures

Cover image courtesy of NASA/CXC/Wisconsin/D.Pooley and CfA/A.Zezas and NASA/CXC/M.Weiss; Figure 1: Science Photo Library; Figure 1.1: Emilio Segrè Visual Archives/American Institute of Physics/Science Photo Library; Figure 1.4: Science Photo Library; Figures 2.2 and 2.5: Copyright © CERN; Figure 2.6: US Department of Energy/Science Photo Library; Figure 2.8: © Corbis; Figure 4.1: Sheila Terry/Science Photo Library; Figure 5.1: Emilio Segrè Visual Archives/American Institute of Physics/Science Photo Library; Figure 6.1: Physics Today Collection/American Institute of Physics/Science Photo Library; Figure 6.2: Los Alamos National Laboratory/Science Photo Library; Figure 6.3: Emilio Segrè Visual Archives/American Institute of Physics/Science Photo Library; Figure 6.18: Time and Life Pictures/Getty Images; Figure 6.21: Corbin O'Grady Studio/Science Photo Library; Figure 7.10: NASA Marshall Space Flight Center (NASA-MSFC); Figure 7.12: ESA, NASA and Felix Mirabel (French Atomic Energy Commission and institute for Astronomy and Space Physics/Conicet of Argentina); Figure 7.13: http://www.star.le.ac.uk; Figure 7.17: NASA Johnson Space Center Collection; Figure 7.18: ESA, NASA, Richard Ellis (Caltech, USA) and Jean-Paul Kneib (Observatoire Midi-Pyrenees); Figure 7.19: Weisberg, J. M. and Taylor, J. H. (2004) 'Relativistic binary pulsar B1913+16: thirty years of observations and analysis', *Binary Radio Pulsars*, Vol TBD, 2004. ASP Conference Series; Figure 7.20: Volker Steger/Science Photo Library; Figure 8.1: M. Collness, Mount Stromlo Observatory; Figure 8.2: Robert Smith, The 2dF Galaxy Redshift Survey; Figure 8.3: Bennett, C. L. et al. 'First Year Wilkinson Microwave Anistropy Probe (WMAP) Observations: Preliminary Maps and Basic Results', *Astrophysical Journal* (submitted); Figure 8.6: NASA Goddard Space Center; Figure 8.11: Ria Novosti/Science Photo Library; Figure 8.13: Adapted from Landsberg, P. T. and Evans, D. A. (1977) *Mathematical Cosmology*, Oxford University Press; Figure 8.15: Sky Publishing Corporation; Figures 8.17 and 8.18: Mark Whittle, University of Virginia from Carr, B. and Ellis, G. (2008) 'Universe or multiverse?', *Astronomy and Geophysics*, Vol. 49, April 2008.

Every effort has been made to contact copyright holders. If any have been inadvertently overlooked the publishers will be pleased to make the necessary arrangements at the first opportunity.