1

Space–time diagrams and the foundations of special relativity

1.1 The concept of a space–time

Space and time are notions familiar to everyone. We shall explore the way in which they form a single entity called *space–time*, firstly according to the ordinary everyday view of how events occur (i.e. according to Newtonian theory). In later sections we shall examine the space–time description of relativity theory.

Space–time according to a single observer

Consider a cine camera set up above a billiard table, pointing directly down to take a series of photographs of the billiard balls on the table (Fig. 1.1a). We may use x and y coordinates to express the position of each of the balls, and could even make these coordinates explicit by marking a coordinate grid on the billiard table. Suppose that one of the balls moves as time progresses, while the rest are stationary. Then the x and y coordinates of this ball will change with time according to this motion, and this will be reflected in the photographs.

Now imagine cutting the cine film to separate the images (Fig. 1.1b) and then stacking these photographs one above the other in their correct time sequence, with the earliest photograph at the bottom and the latest at the top (Fig. 1.1c). The position of each ball at any time $t = t'$ is represented by the position of its image in the corresponding photograph, with its successive positions at later times recorded in the subsequent photographs higher up in the stack. Thus a glance at this stack of pictures will show the way the arrangement of the balls changes with time; in particular it will show how one ball moves and the others are all stationary.

This stack of photographs already contains the essential idea of a space–time, namely the presentation of a time sequence of images one above the other showing the successive positions of objects in the space considered (here, the surface of the table). However, there is one problem: the stack of photographs is very frail: one sneeze will destroy its order. To remedy this, imagine taking the stack of photographs and fusing them together in an oven, to obtain a solid, durable space–time (Fig. 1.1d). This is a three-dimensional space–time, with the vertical axis

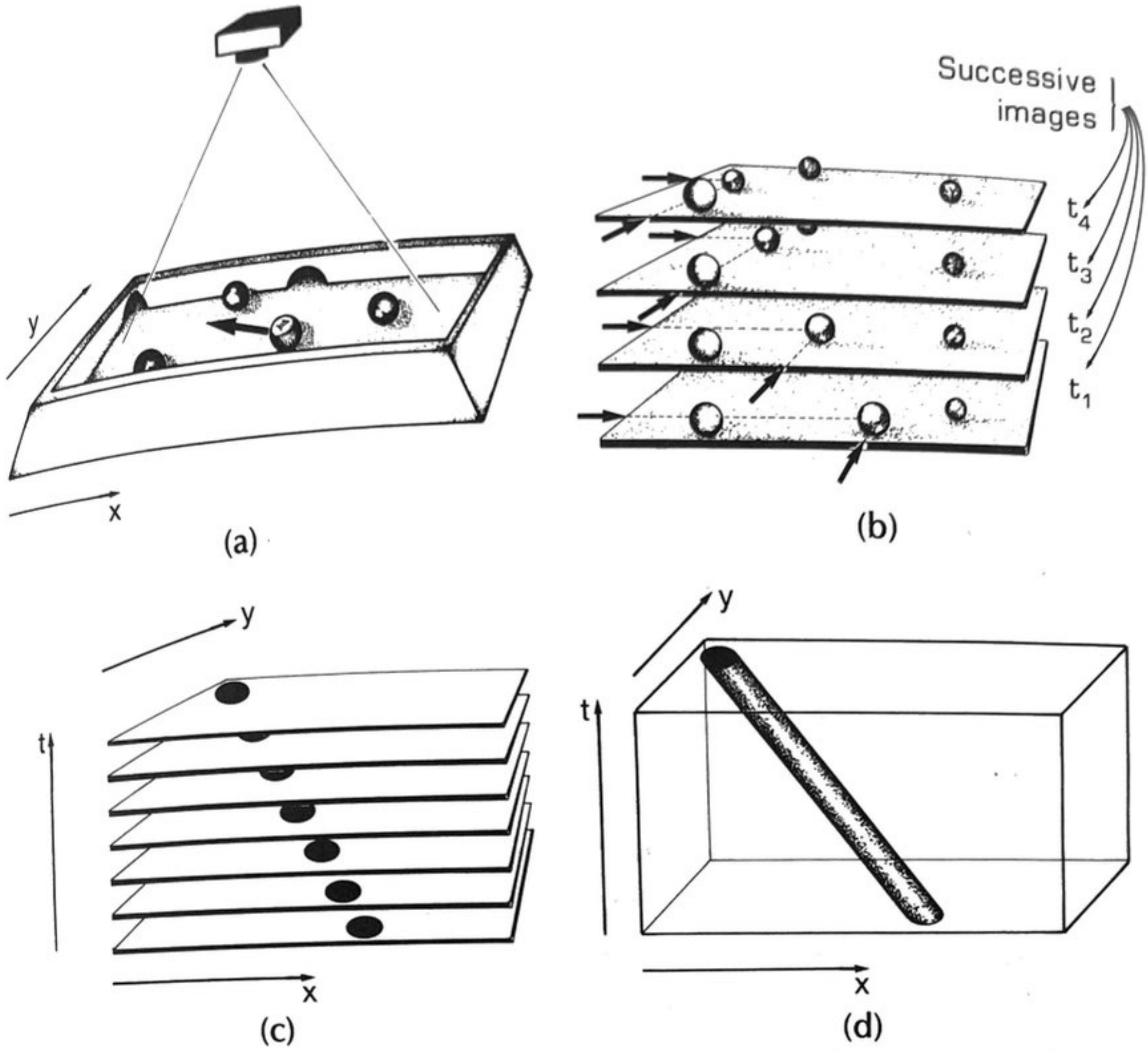


Fig. 1.1 Constructing a space–time. (a) A cine camera takes photographs of billiard balls on a table. One ball moves relative to the others. (b) A series of photographs from the film. (c) The photographs stacked together, later ones above the earlier ones. (d) The photographs fused together to form a 'space–time', with time coordinate t and spatial coordinates x, y.

depicting time, represented by a coordinate t (measured by a clock), and the horizontal axes depicting spatial position on the surface of the table, represented by coordinates x and y (measured by rulers). The space–time represents the histories of all objects in the two-dimensional space. Thus the histories of the stationary billiard balls are represented by vertical tubes in the space–time, while the history of a ball moving to the left is represented by a tube sloping over to the left. To recover the detailed history of motions of objects in the space, simply consider a series of horizontal sections of the space–time (surfaces of instantaneity) at later and later times. These sections intersect the tubes representing the histories of the stationary balls at x and y coordinate positions that stay constant (showing that they are indeed stationary), and intersect the tube representing the ball moving to the left in positions that are successively more to the left (showing it does indeed move to the left). In effect, by considering a succession of time slices in this way one can reconstruct a series of images corresponding to the photographs from which the space–time was initially constructed, and then by considering these in turn one can visualize the motion of the particles as in a cine film. The space–time therefore completely represents these motions.

The space–time we have constructed is three-dimensional, representing the histories of objects in a two-dimensional space (the surface of the table). Of course, real space–time is four-dimensional, with three space dimensions (described by coordinates x, y, z) and one time dimension (described by the coordinate t), representing the histories of all objects in three-dimensional space. We cannot easily represent this in a single picture. However a study of three-dimensional (or even two-dimensional) space–times will enable us to understand many of the properties of the full four-dimensional space–time. We will demonstrate this in the rest of this book.

Space–time according to different observers

Different observers will in general have different views of the space–time. Returning to consider the billiard table discussed above, we suppose now that in addition to a camera A held fixed above the billiard table, (Fig. 1.2a), there is a second camera B, which moves with the moving ball (Fig. 1.2b).* To simplify matters suppose that the ball moves parallel to the x-axis; then the camera will also move parallel to the x axis at the same speed as the ball, directly above it, so that the ball stays at a fixed position in the viewfinder. Then in the space–time model constructed from the pictures obtained by A (exactly as described above) the history of the moving ball is a tube slanted to the left (Fig. 1.2c), while in the space–time model constructed by B (again, exactly as above) the history of this ball is a vertical tube (Fig. 1.2d). This is because the ball moves to the left relative to the coordinate x corresponding to A's view, but stays fixed in the coordinate x' corresponding to B's view. Thus we have two different views of the same set of happenings. These are the same space–time described from different viewpoints.

This illustrates one of the major issues that arises in understanding space–times: one can use different coordinate systems, corresponding to making different sets of observations, to study the same physical system. The space–time representations arising will apparently be different, but can in fact be transformed into each other by making the appropriate changes of coordinates. Later we will determine the mathematical transformations that relate the viewpoints of the two observers. For the present, we simply note that when we consider the series of photographs from which the space–time representations are constructed, the relation is a simple one. Suppose that before we fuse A's set of photographs

* If you feel that the labels A and B for the different cameras and the corresponding observers are antiseptically impersonal, you might like to substitute names such as Alfred or Angela for A, Barbara or Bernard for B. While such labelling may well initially help the beginner to grasp what is happening, ultimately it becomes an annoying distraction. We have chosen to use the more convenient abstract labels from the beginning.

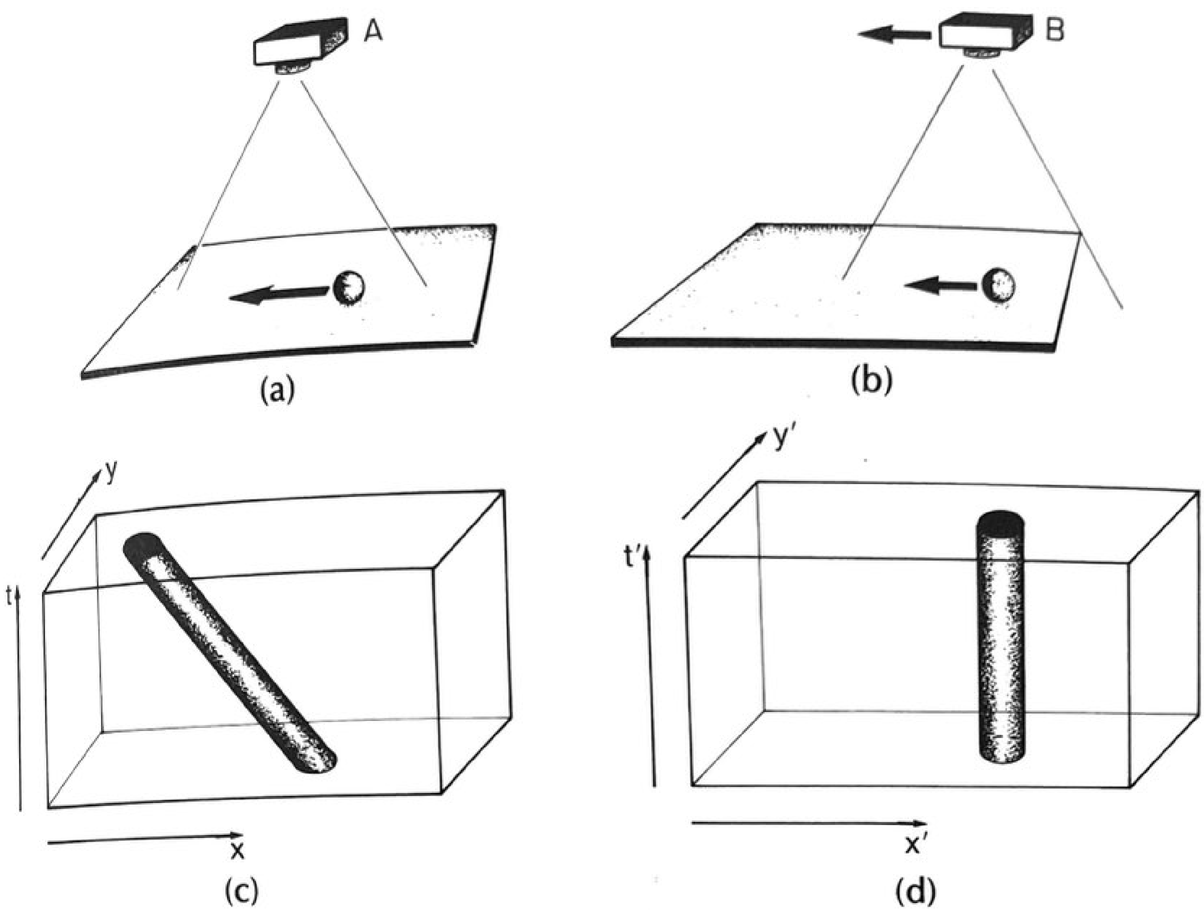


Fig. 1.2 Effect of the observer's motion on the space–time picture. (a) Camera A is fixed above the billard table. (b) Camera B moves with the moving billiard ball. (c) The space–time view of the ball's history, constructed from A's photographs. (d) The space–time constructed from B's photographs.

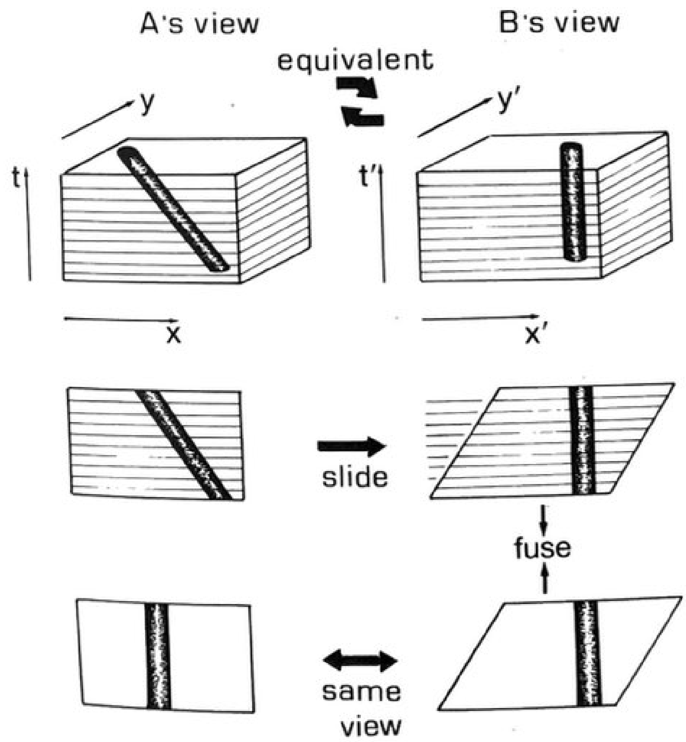


Fig. 1.3 Although they look different, A's and B's space–time views are equivalent: sliding A's pictures sideways before fusing them together will give the same space–time view as B's.

together, we slide them carefully sideways until the images of the moving ball are directly above each other (Fig. 1.3); then A's and B's representation of the same set of physical events will be the same. By this means, the view obtained by the first camera has been transformed into the same as that obtained by the second.

Examples of space–times

The ideas explained so far should become quite clear on carefully considering two examples.

(A) A planet in circular motion around a sun. In the sun's frame of reference, the sun is at rest in the spatial coordinates used, while the planet circles around it, describing a helix in space–time (Fig. 1.4). To see that this is the correct space–time picture, consider later and later time sections of the space–time; the positions of the planet in the successive surfaces of instantaneity trace out a circle around the sun, as required.

(B) A circular wave in a pond. Consider dropping a stone into a large pond at some time t_1, producing a spreading spherical ripple in the pond (Fig. 1.5a). Photographs of the crest of the spherically spreading wave taken from a camera stationary above the point of impact (Fig. 1.5b) produce a space–time picture in which the spreading wave is depicted as a cone with apex at time $t = t_1$ (Fig. 1.5c). Again considering later and later surfaces of instantaneity in the space–time, we recover the series of images depicting the spherically spreading wave, starting from the centre at time t_1.

Points in space–time are called *events*. An event represents a particular position in the physical world at a particular time, the set of all events representing the spatial and temporal locations of all possible physical occurrences. A *world-line* is the path traced out in space–time by the events representing the history of a particular particle or light ray. For example the helix in example (A) is the world-line of the planet as it orbits around the sun. Not all lines in space–time are possible world-

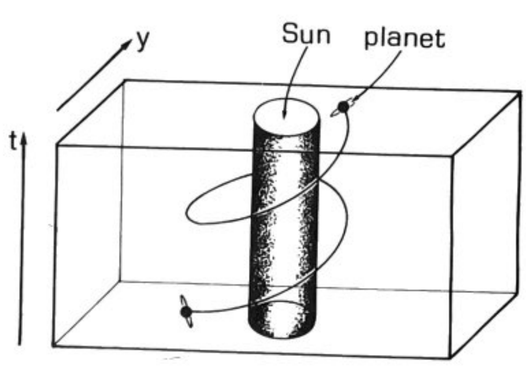


Fig. 1.4 A planet in circular motion around the Sun, describing a helix in space–time.

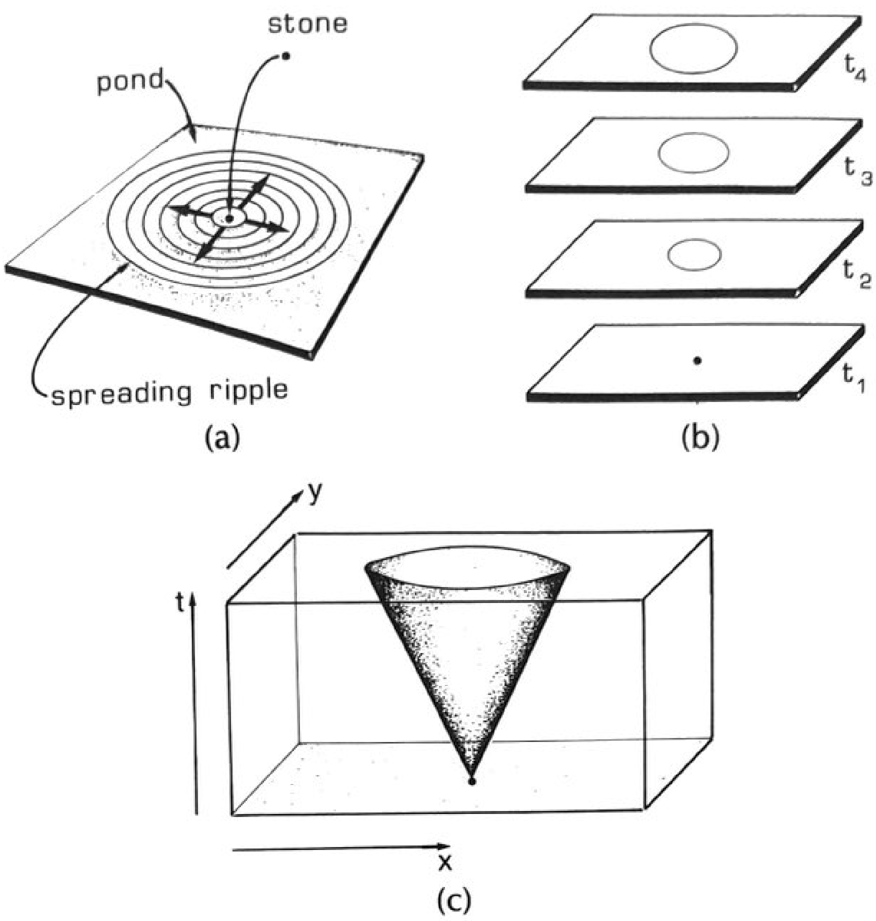


Fig. 1.5 (a) Circular ripples produced by a stone thrown into a pond. (b) A succession of photographs of the spreading wave. (c) A space–time view of the spreading wave.

lines; for example, if a line reaches a maximum time and then slopes down again (Fig. 1.6), it does not represent a possible world-line of a massive body, because time would start to go backwards along such a world-line, where it slopes down. We shall discover further restrictions on allowable world lines after considering the limiting role played by the speed of light in relativity.

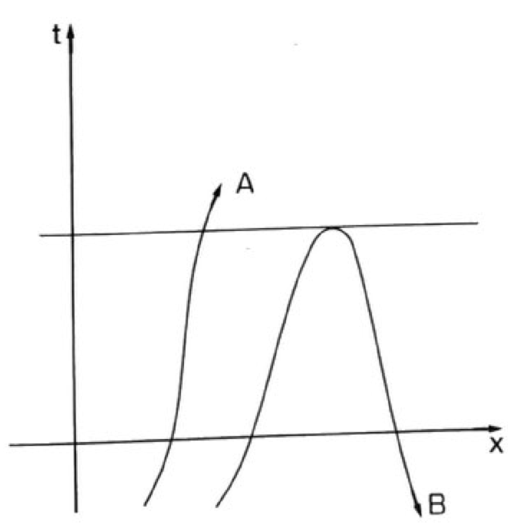


Fig. 1.6 Curves in space–time: A is a possible particle history, or *world-line*; B is not.

Summary

Space–time represents the histories of objects in space. When the space represented is two-dimensional, the space–time is three-dimensional (three coordinates are needed to characterize all events: the two spatial coordinates x and y depicting the spatial position of the event, and the coordinate t representing the time of the event). The full space–time needed to represent all events in the real physical world is four-dimensional (with one time coordinate and three spatial coordinates). Each surface ($t =$ constant) tells us where each object was at the time t, according to an observer using a particular coordinate system, say (x, y, z); these surfaces are slices of instantaneity or simultaneity in the space–time (Fig. 1.7).

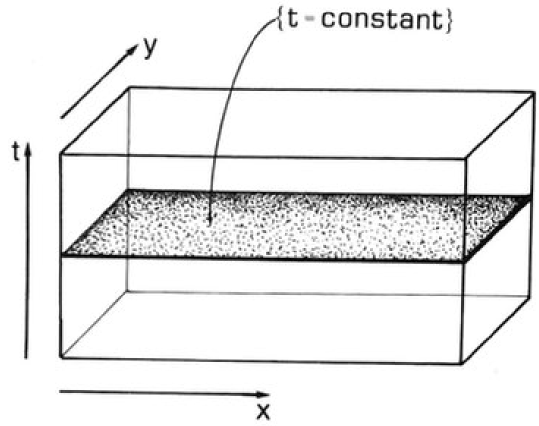


Fig. 1.7 A $\{t = \text{constant}\}$ slice of a space–time; this represents a surface of simultaneity.

Exercises

1.1 An observer O watches the engine of a train shunting on a straight track; he chooses the x coordinate to measure distance along the track. Plot the world-line of the engine in the (t, x) plane if, starting at a distance of 50 m from the observer, (i) it moves at 10 m/sec away from the observer for 5 seconds; (ii) then it is stationary for 7 seconds; (iii) then it moves at 5 m/sec towards the observer for 8 seconds.

1.2 The motion of a rocket relative to observer A is shown in Fig. 1.8. What

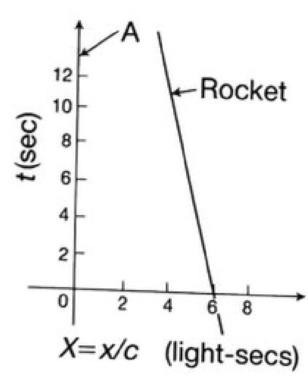


Fig. 1.8

is the distance of the rocket from A at $t = 0$ years? at $t = 10$ years? What is the speed of motion of the rocket relative to A?

1.3 Draw a space–time diagram representing the motion of the Moon about the Earth (stating carefully what reference frame you are using). Indicate approximate time and spatial scales on your diagram.

1.4 Suppose a particle in an accelerator moves in a circular orbit of radius 25 m, speeding up all the time as it moves. Sketch a space–time diagram of its motion.

1.5 Two cars A and B, watched by a person C waiting to cross the street, collide and then bounce apart. Sketch the world-lines of A, B, and C as seen by (i) the driver of one of the cars; (ii) the driver of the other car; (iii) the person waiting to cross the street. [The drivers are each securely seat-belted into their respective cars.]

So far, our discussion of space–times has been based on the everyday ideas of Newtonian theory. The concept of a space–time applies equally in the case of relativity theory, provided we take into account important relativity principles which we examine in the next two sections.

1.2 Causality and the speed of light

The speed at which light travels is very large but nevertheless is finite. It is measured to be approximately 3×10^{10} cm/sec $= 3 \times 10^8$ m/sec $=$ 300 000 km/sec. Thus, for example, light travels 30 km in 10^{-4} sec $=$ (1/10 000) sec, and 300 km in 10^{-3} sec $=$ (1/1000) sec. According to the Newtonian view of space–time, there is nothing special about the speed of light, and physical influences (e.g. changes in a gravitational field) can propagate faster: indeed, in principle they can influence distant regions instantaneously. According to relativity theory, the situation is quite different.

The limiting nature of the speed of light

One of the basic principles of Einstein's special theory of relativity is that the speed of light is a limiting speed for all communication and for all motion of massive bodies; indeed it is a limiting speed for propagation of all causal influences. One should note here that this speed is the speed of travel of all electromagnetic radiation, not merely light; it is the speed of travel of infrared and ultraviolet radiation, of radio waves and X-rays, as well as visible light (because these are all forms of electromagnetic radiation, at different wavelengths). Further, it will be the speed of travel of any particles of zero rest mass there may be, e.g. gravitons and massless neutrinos as well as photons. Thus one can send signals at the speed of light in many ways, but there is no way one can send a signal faster. Any massive object, e.g. a rocket, a meteorite, a human being, cannot travel as fast as light.

There is experimental evidence for this principle from many sources. On the one hand, no particle or signal has ever been measured to move faster than this speed. On the other, attempts to accelerate objects to higher speeds fail. For example, suppose one accelerates particles in a linear accelerator, and then plots the square of the resulting speed against the energy given to the particles. Newtonian theory predicts that no matter how high the speed, the resulting graph will be a straight line because the kinetic energy of the particle is proportional to the square of its speed of motion; in particular, there should be no barrier to accelerating particles to move faster than light. In practice it turns out that the Newtonian prediction is correct at low speeds, but at higher speeds the experimental results deviate from this prediction: the speed attained is less than that predicted by Newtonian theory. This happens in

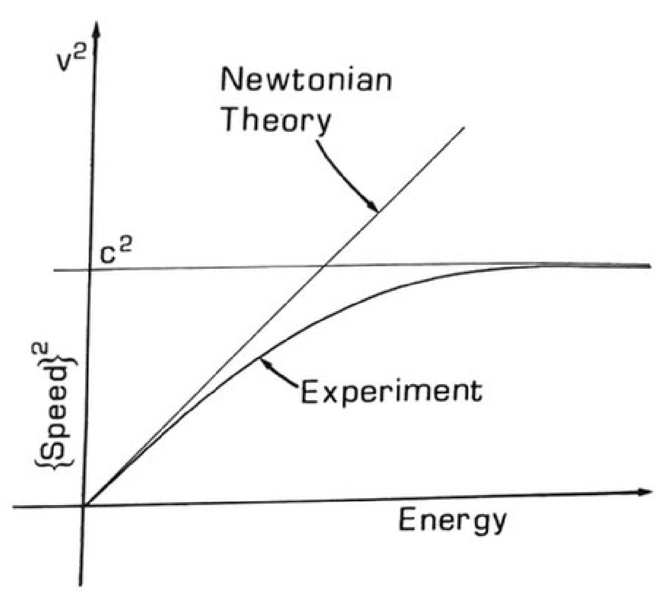


Fig. 1.9 A graph of the square of the speed of a particle against the energy of motion given to it, showing the experimental result and the prediction of Newtonian theory. No matter how much energy is given to the particle, the speed of light *c* is a limit to the speed it attains.

such a way that no matter how much energy one imparts it is not possible to accelerate particles to move faster than the speed of light (Fig. 1.9). The amount of energy needed to accelerate fast-moving particles to higher speeds becomes larger and larger as the speed increases; smaller and smaller speed increments result from each doubling of the energy, and the speed of light is never reached. This is an experimental result that has been proved many times over at a cost of many billions of dollars (since that is the cost of the high energy particle accelerators now in use). One has to invest large sums of money in accelerators to produce an observable effect, because the speed of light is so large: the speed-of-light limit certainly does not act as a factor restricting the speed of cars, aircraft, or other vehicles on the earth!

The need to allow for the speed of light

The time delay between lightning and thunder reminds us to allow for the speed of sound, but that is not the only allowance we should make! The limiting nature of the speed of light in special relativity means that one should always allow for light travel time in analysing any physical phenomenon. As an example, any photograph will, in general, include images of objects at various distances and so various light travel times. This means the images in a photograph will represent the states of the objects pictured at different times in the past. Thus a photograph of the Moon framed by trees represents the state of the moon 1.27 seconds earlier than that of the trees; a photograph of distant galaxies with foreground stars (Fig. 1.10) represents delays of millions of years in the state of the galaxies relative to the stars (the stars will typically be at distances for which the light travel time is thousands of years but the galaxies at distances for which the light travel time is millions of years). In each case we see the object at the instant where the light was emitted; the camera therefore necessarily records the resulting time delays. The cover photograph of this book graphically shows a time delay of 8 minutes 18.7 seconds (the time taken for light to travel from the Sun to the Earth), because the image of the Sun represents the situation there about eight minutes before the Earth was illuminated by that light. Because the camera records images by means of light arriving at it at one instant, the photograph depicts the state of the Sun eight minutes earlier than the state of the Earth pictured in the same image.

To explore this effect further, consider a camera 3 metres above the centre of a circular pond of diameter 8 metres (Fig. 1.11a). The light has to travel a distance of 3 metres from the centre of the pond to the camera, taking $(3\,\text{m})/(3 \times 10^8\,\text{m/sec}) = 10^{-8}$ seconds to do so, but light from the edge of the pond has to travel a distance of 5 metres, taking $(5\,\text{m})/(3 \times 10^8\,\text{m/sec}) = \frac{5}{3} \times 10^{-8}$ seconds to do so. Thus light from the edge takes $\frac{2}{3} \times 10^{-8}$ seconds more to reach the camera than light from the centre. A photograph records one instant when light reaches the camera from different places within its field of view; if these places are at various distances from the camera, the image obtained will represent the different times when the light set out towards the camera. Hence, when the camera takes a photograph of the pond, one will obtain images of the situation in different areas of the pond at different times: light from the edge has to travel further and so has to set out earlier in order to reach the lens at the same time as light from the centre. If we sketch lines of exact simultaneity on a photo P_1 of the pond taken by the camera, they will form circles with the outer circle depicting the situation at the pond earliest, say at a time t_1, and the central point the situation at a time t_2 which is 0.667×10^{-8} seconds later than t_1 (Fig. 1.11b). A photograph taken by the camera is not an instantaneous photograph of the pond!

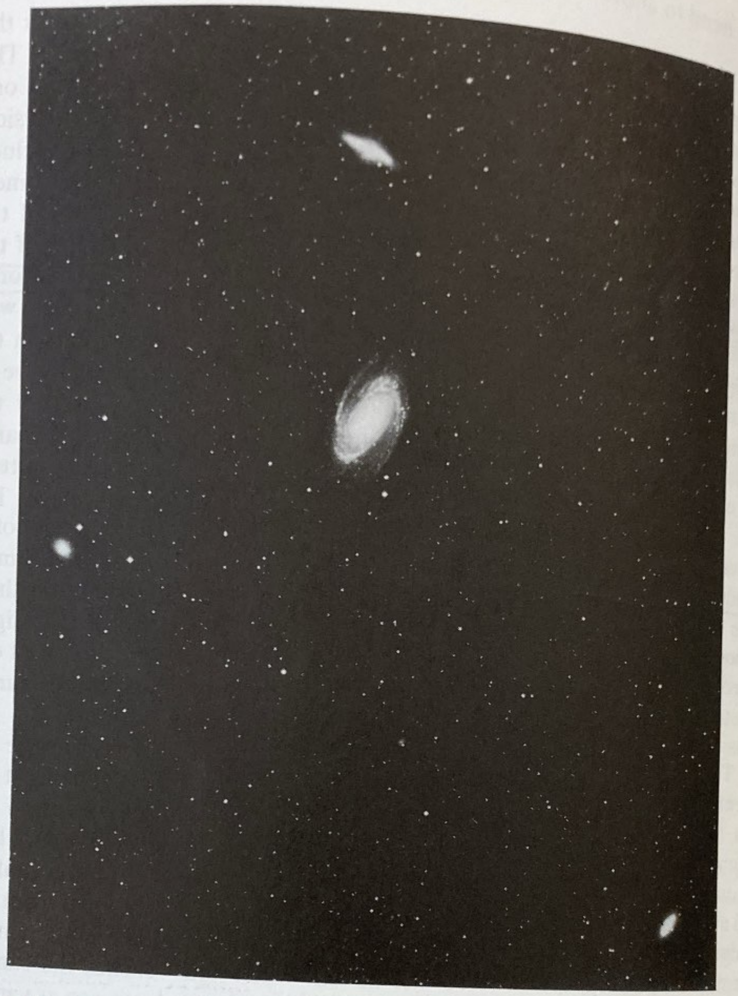


Fig. 1.10 Distant galaxies and foreground stars. The foreground stars all belong to our own galaxy, which is a spiral system of stars and dust like the galaxy M81 shown here. The four 'nearby' galaxies visible in the photograph are at a distance of some millions of light years from us (three fainter galaxies are even more distant) but the individual stars seen are within a few thousand light years. The photograph dramatically illustrates the time delays necessarily involved in all our observations of distant objects: we are seeing conditions at the galaxies millions of years ago, and those in the stars up to a few thousand years ago. Thus the images represent these objects as they were at times differing by millions of years. (Photograph from the Hale Observatory.)

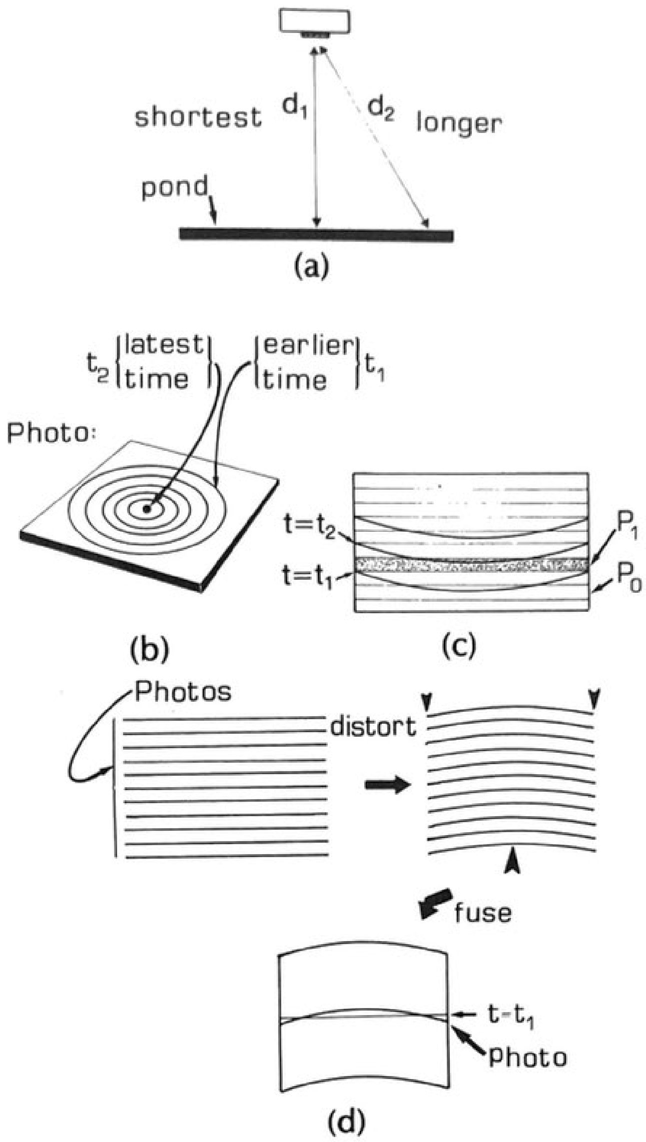


Fig. 1.11 (a) A camera above the centre of a pond: the distance d_1 to the centre is clearly shorter than the distance d_2 to a point further out. Consequently, light arriving at the camera from the centre set out later than light arriving at the same instant from the edge. (b) Circles of constant imaging time on a photograph P_1 of the pond, the larger circles corresponding to earlier times. (c) Surfaces of simultaneity in a stack of photographs of the pond (viewed edge-on, showing the finite thickness of each photograph). The photograph P_1 is shown shaded. (d) Distortion of the stack of photographs before fusing, to represent correctly surfaces of simultaneity as exactly horizontal sections of space–time.

Hence, on stacking a succession of photographs together and fusing them to obtain a representation of space–time, horizontal sections will not represent exact simultaneity:* as one moves out from the centre on a horizontal slice of space–time (which will be one of the photographic images), the situation represented will be earlier and earlier the further one is from the centre. There will be an earlier photo P_0 in which the

* In Section 1.1, we ignored light travel time and so regarded horizontal slices as exactly simultaneous. This will be a good approximation for slowly moving objects considered at everyday time and length scales.

situation at the central point is depicted at the time t_1; this photograph will lie below P_1 in the stack (because later photographs lie above earlier ones). It follows that exact surfaces of simultaneity in the space–time (e.g. t_1 is constant) will be lowest at the centre and will curve up as one moves from the centre to the edge (Fig. 1.11c).

To correct this, i.e. to obtain a space–time representation in which horizontal sections are indeed exactly simultaneous sections of the space–time, one will have to distort the photographs of the pond by bending their outer regions downwards before stacking them and fusing them together (Fig. 1.11d). One could in this way allow for the light travel time, and obtain a space–time picture correctly representing simultaneity as exactly horizontal surfaces.

In this particular case, the effect is negligible in practice. However, this will not always be true. Consider, for example, the delays implied from the centre to the edge of the photographic image where an observer in a spacecraft photographs the disc of a galaxy from a distance of 30 000 light years above the centre of the galaxy. If the galaxy has a radius of 40 000 light years, the delay represented in the photograph will be 20 000 years, i.e. the situation at the centre will be depicted 20 000 years after that at the edge of the disc.

Light rays in space–time

In flat space, light travels in straight lines; as it travels at the constant speed c, the path traced out in space–time by light (strictly, by a *photon*, that is, a light particle) will also be a straight line. Each light ray in space–time represents travelling a distance d in a time t given by $t = d/c$, where the symbol c is used to represent the speed of light (so $c = 3 \times 10^{10}$ cm/sec). For example, if a light ray is emitted in the x direction at the event O with coordinate values $x = y = z = 0$ with $t = 0$, then in 1 second it will be at the position $x = 1c$ cm $= 3 \times 10^{10}$ cm with $y = z = 0$; at the time $t = 2$ seconds, it will be at the position $x = 2c$ cm $= 6 \times 10^{10}$ cm with $y = z = 0$; and so on (Fig. 1.12a). It is convenient to

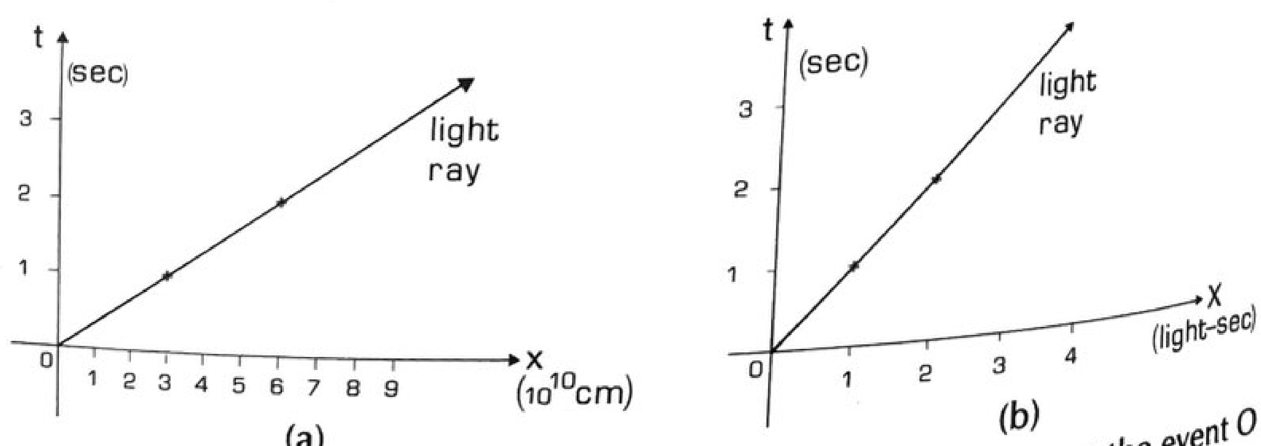


Fig. 1.12 (a) A light ray travelling in x-direction after emission at the event O ($x = 0$, $t = 0$). Its space-time position is shown at $t = 1$ and $t = 2$. (b) The same light ray depicted using a spatial coordinate $X = x/c$ (with units of light-seconds).

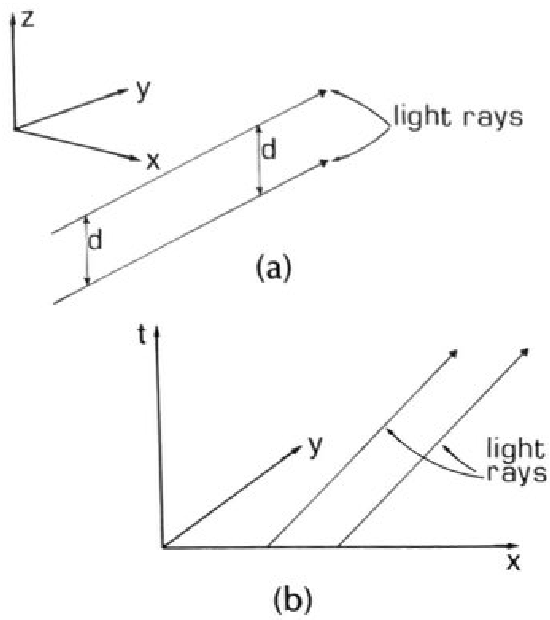


Fig. 1.13 (a) Parallel light rays in a three-space with coordinates (x, y, z). (b) These rays are represented by parallel straight lines in space–time.

measure spatial distances in terms of coordinates $X = x/c$, $Y = y/c$, $Z = z/c$ which are just the previous spatial coordinates divided by the speed of light; they are the same distances but measured in terms of 'light-times' (light-seconds, light-years, etc.). Then in 1 second the light would be at the position $x = 1c$ cm, $y = z = 0$, so $X = (1c \text{ cm})/(c \text{ cm/sec}) = 1$ light-second, $Y = Z = 0$; at the time $t = 2$ seconds, it will be at the position $X = (2c \text{ cm})/(c \text{ cm/sec}) = 2$ light-seconds, $Y = Z = 0$; and so on. At an arbitrary time t, it will be at the position $X = (ct)/c = t$ light-sec, $Y = Z = 0$ (Fig. 1.12b). The relation between this and the previous representation is easily obtained on remembering that 1 light-second $= (1 \text{ sec}) \times (c \text{ cm/sec}) = 3 \times 10^{10}$ cm $= 300\,000$ km. Another way of thinking of the coordinates X, Y, Z is that when they are used, we have effectively chosen units of measurement for spatial distances so that the speed of light is 1 (because then light travels a distance of 1 light-second in 1 second, etc).

In flat space, initially parallel light rays never meet each other because the spatial distance between them stays constant (Fig. 1.13a); consequently in space–time diagrams, they are represented by parallel straight lines that remain a constant distance apart (Fig. 1.13b). We shall see later that this is not true in a curved space–time.

The light cone and causal regions

The *future light cone* of an event O is the set of all light rays through that event (Fig. 1.14). This represents the space–time paths of light rays emitted in all directions from that place and time. It may conveniently be thought of as the history in space–time of a flash of light emitted in all directions at the position and instant corresponding to the event O; thus one can imagine a flash bulb going off at this place and time, resulting in

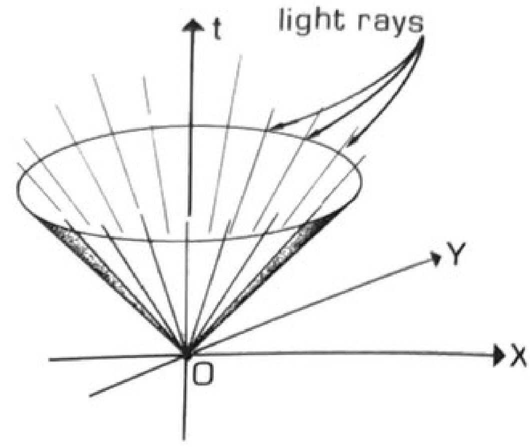


Fig. 1.14 The future light cone of the event O is the set of all future-directed light rays through O.

a sphere of light spreading out in all directions at the speed of light. At a time t after the flash was emitted, the light forms a sphere at distance $d = ct$ from the source position (Fig. 1.15a). For definiteness, let us assume the event O is $(x = y = z = 0,\ t = 0)$.

It is difficult to represent the full light cone in a diagram, so we restrict our attention to a fixed value of z, say $z = 0$, obtaining the projection of this spreading light in a two-dimensional plane. The light will spread out circularly in this plane, which is described by coordinates x and y. This is exactly analogous to the spherical wave in the pond (Example (B) above). By exactly the same reasoning as used in that example (leading to Fig. 1.5c), a three-dimensional space-time diagram representing the spread of the light will show the wave front as a cone originating at $(x = y = 0,\ t = 0)$ and with radius ct at time t (Fig. 1.15b). As the future light cone of the event O obtained in this way represents light travelling out in all directions from the emission event O, it is generated by all the future light rays that pass through O.

To represent this situation in a clear, standard way, it is convenient to use the coordinates $X = x/c,\ Y = y/c,\ Z = z/c$ introduced above. Their

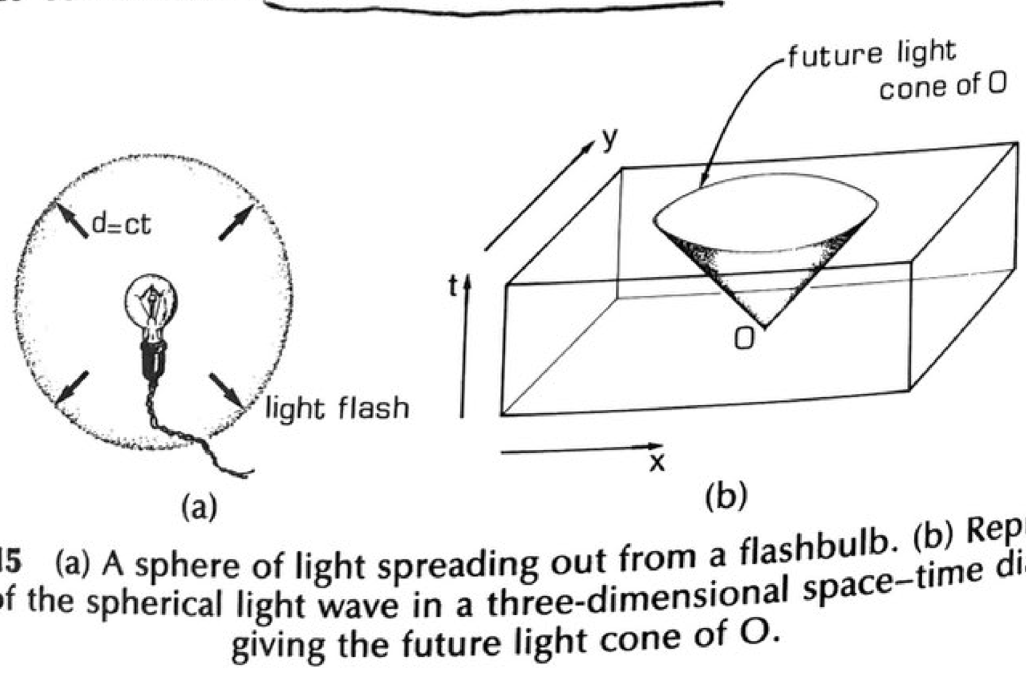


Fig. 1.15 (a) A sphere of light spreading out from a flashbulb. (b) Representation of the spherical light wave in a three-dimensional space–time diagram, giving the future light cone of O.

use has the advantage that in these units the spatial distance travelled is equal to the time elapsed (the effective speed of light is 1); for example, after a time of 1 second, the light has spread to a sphere of radius 1 light-second. Consequently the light cone makes an angle of 45° with the vertical axis, representing the fact that a unit horizontal distance in these diagrams is traversed in a unit time; this makes it particularly easy to draw the light cones when these units are used (Fig. 1.15b was drawn using this convention).

It is often convenient to restrict our attention even further to a fixed value of Y (say $Y=0$) as well as a fixed value of Z. The light then spreads out in a one-dimensional space with X as the spatial coordinate (this situation might be realized, for example, if a pair of optical fibres convey the light from the flashbulb in the positive and negative X directions, Fig. 1.16a). The corresponding two-dimensional space–time diagram shows the light emitted from the event O as travelling on lines at ±45° to the t axis (Fig. 1.16b); these are the two light rays through O, because such lines are precisely those in which a unit (vertical) change in time corresponds to a unit (horizontal) change in distance. This diagram is a two-dimensional section (with one time and one space dimension represented) of the three-dimensional Fig. 1.15b (representing one time

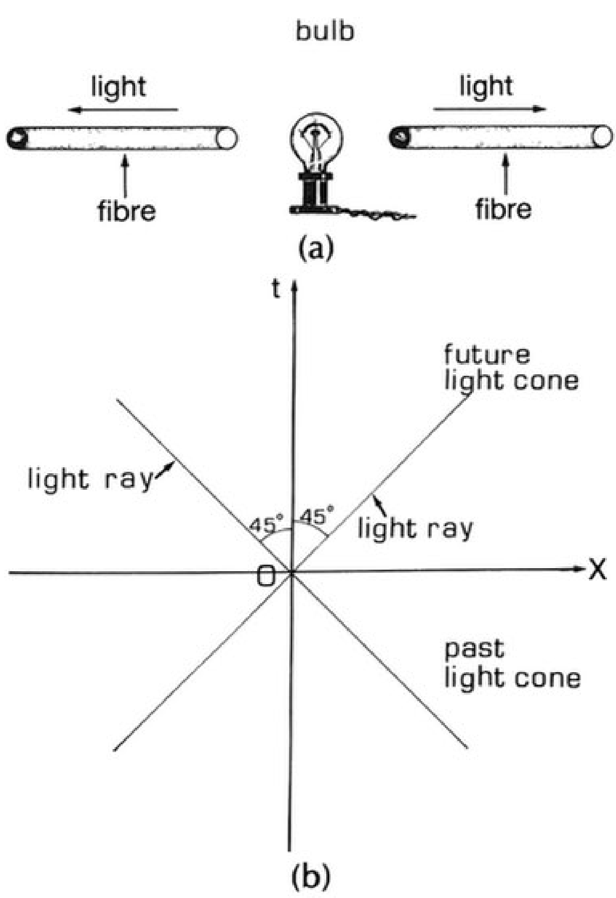
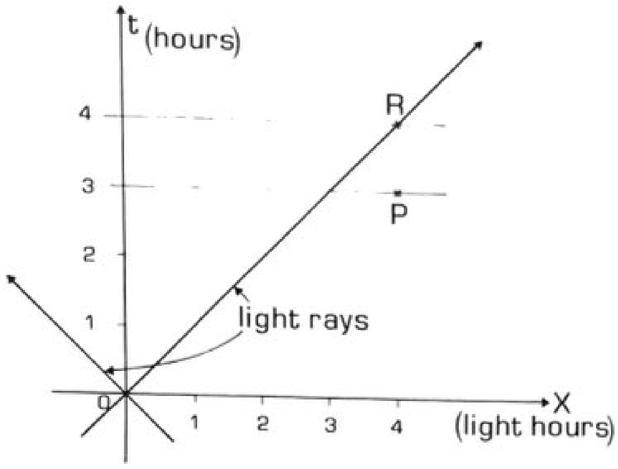


Fig. 1.16 (a) Light spreading from a flashbulb one-dimensionally along optical fibres. (b) Representation of these light rays in a two-dimensional space–time diagram, generating the future light cone of O. The past light cone of O (i.e. light rays converging to O) is also shown.

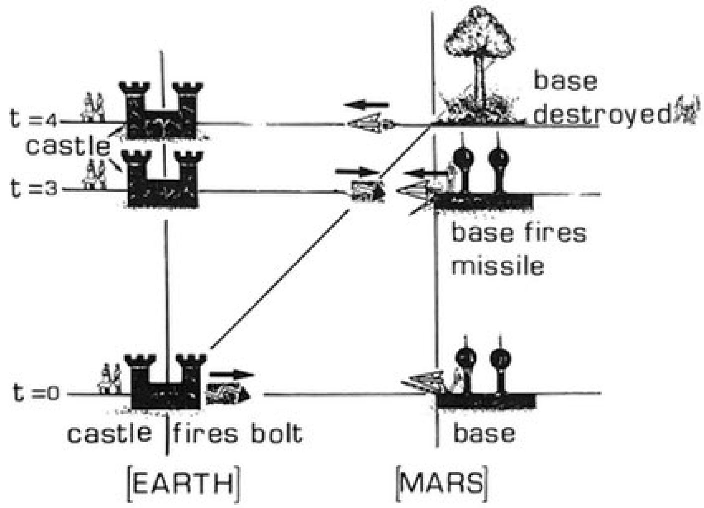
and two space dimensions). In this diagram we have extended the light rays to the past of O; the light rays converging on O from the past generate its *past light cone,* representing converging light pulses that arrive at the position ($X=Y=Z=0$) at the time $t=0$.

The importance of the light cone of any event derives from the fact that it limits the region of space–time which can be causally affected from that event. For example, suppose President Lugarnev of Transylvania receives information at noon that at 3:00 p.m. a nuclear missile is to be launched towards his castle on the earth from a secret base on Mars. He instantly presses the button firing his Super-Z lasers at the base on Mars, but he is too late: the energy bolts he has released, travelling at the speed of light, will take 4 hours to reach Mars and so will destroy the rocket launching pad 1 hour after the missile has left. Let the event where he receives the information be O; this event (specified by a time and spatial position) is then noon at his castle. The light cone of O is depicted in Fig. 1.17, where, for convenience, time is measured in hours from O and spatial distances in light-hours from O (so O has the coordinates $t=0$, $X=0$). Then the event where the missiles are to be launched is P, given by $t=3$, $X=4$. The light cone clearly shows that the laser beam emitted at O will arrive at Mars too late to influence P. One cannot influence P from O, because it is outside O's light cone.

The reason for this limitation, of course, is the limiting nature of the speed of light. The angle of a particle's world-line in space–time from the vertical depends on rate of change along the world-line of spatial distance with respect to time, and so represents the speed of motion of the particle relative to the chosen coordinate system (Fig. 1.18). Therefore, the limiting nature of the speed of light means that no world-line can make a greater angle with the vertical than the light cone; using the coordinates (X, Y, Z), no world-line can make an angle larger than 45° with the vertical axis. Further, one can only send light or radio signals from any event to events on its future light cone. Considering this, it becomes clear that an observer at an event O cannot influence any event that lies outside the future light cone of O (to do so would involve causally influencing events along paths representing motion at speeds greater than the speed of light). This is a fundamental limitation on all communication, implied by special relativity theory. It follows that given any event P, we may divide space–time into five distinct causal regions (Fig. 1.19). The interior of the future light cone $C^+(P)$ is that region which can be influenced by objects travelling from the event P at less than the speed of light; the future light cone itself can be influenced from P by signals travelling at the speed of light. The past light cone represents the set of events in space–time from which signals sent at the speed of light arrive at the spatial position and time represented by event P. Thus in a photograph of an object taken at P, the light arriving at P records the situation at the instant where the object's world-line intersects our past



(a)



(b)

Fig. 1.17 (a) A space–time diagram showing the event P ($t=3$, $X=4$) where missiles are launched from Mars towards the Earth. At the time $t=0$ on the earth (at $X=0$), it is already too late to prevent the launching of these missiles; this is because a laser pulse emitted at this event O will reach Mars at the event R ($t=4$, $X=4$), an hour after the missiles were launched. (b) Depiction of this series of events by a sequence of instantaneous spatial views. At $t=0$, the castle fires a bolt towards the missile base; at $t=3$, the base fires a missile while the bolt is still a light-hour away from it; at $t=4$, the base is destroyed but the missile is on its way to the castle. Note the direct correspondence between these spatial views and the space–time diagram. The reason event P cannot be influenced from event O is because P is outside O's future light cone (the light ray OR lies on this light cone).

light cone (Fig. 1.20); the camera necessarily records the resulting time delays (as in the cover photograph). The interior of the past light cone $C^-(P)$ is the region in space–time from which the event P can be influenced by objects travelling at less than the speed of light. The exterior of the light cones is the region which cannot be influenced by P and which cannot influence P.

One can illustrate the latter feature by considering a particular event on the surface of the Earth, when an astronaut on the Moon is observed

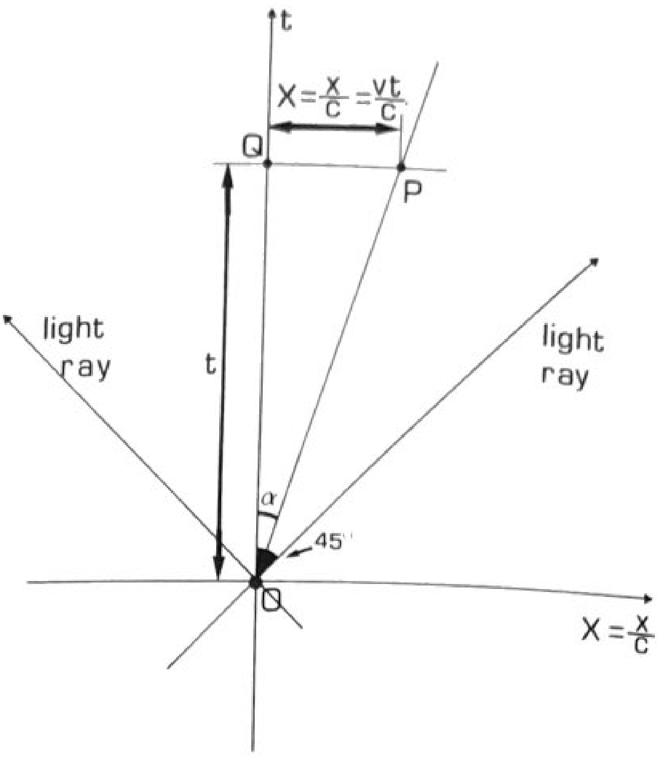


Fig. 1.18 A straight world-line passing through O and P represents motion relative to the reference frame (t, X) at a speed v in the X-direction; at time t, it is at position $X = x/c = vt/c$. The angle α of this world-line to the vertical is given by $\tan\alpha = X/t = v/c$. For a light ray, $v = c$ and $\tan\alpha = 1$.

through an ultrapowerful telescope. Suppose that at this time one were to observe a boulder rolling down a slope towards the astronaut. Since light takes 1.27 seconds to reach the Earth from the Moon, we are observing an event 1.27 light-seconds away and 1.27 light-seconds to the past, on the past light cone (Fig. 1.21). It is already too late to radio a warning to

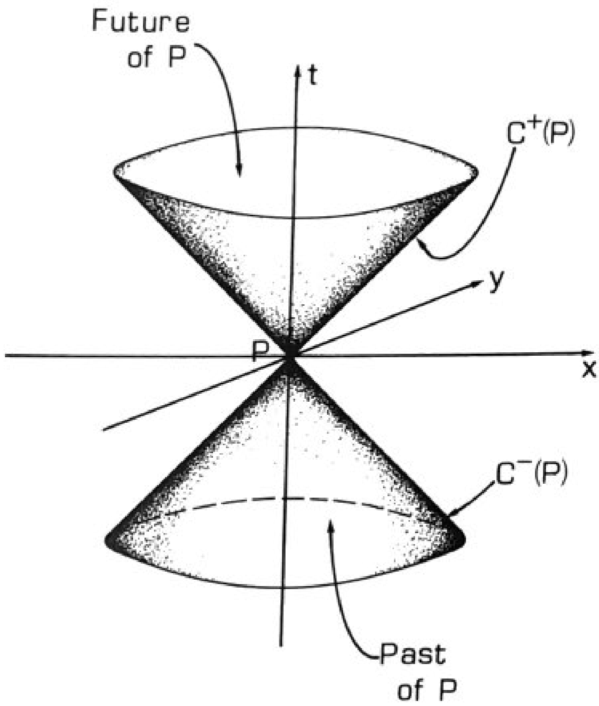


Fig. 1.19 The future and past light cones $C^+(P)$, $C^-(P)$ of an event P determine the *future of* P (the interior of the future light cone), and the *past of* P (the interior of the past light cone). Events outside these light cones cannot be influenced from P or influence what happens there.

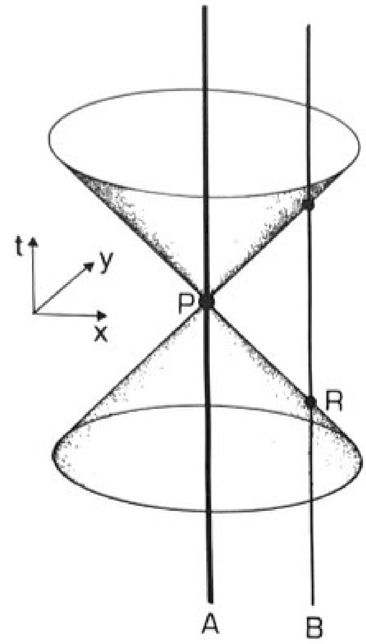


Fig. 1.20 A photograph taken by observer A at the event P depicts the event R in B's history, where B's world-line intersects the past light cone of P.

the astronaut if the boulder will take 2 seconds to reach him, because the event where the bolder will reach him is outside the causal future of the reception event. Given the restrictions on communication resulting from the limiting nature of the speed of light, there is no method of sending a warning signal in time.

The causal limitations discussed here are fundamental, but will not significantly affect ordinary everyday life in an obvious way because the speed of light is so large: in the context of cars, aircraft, etc. on or near

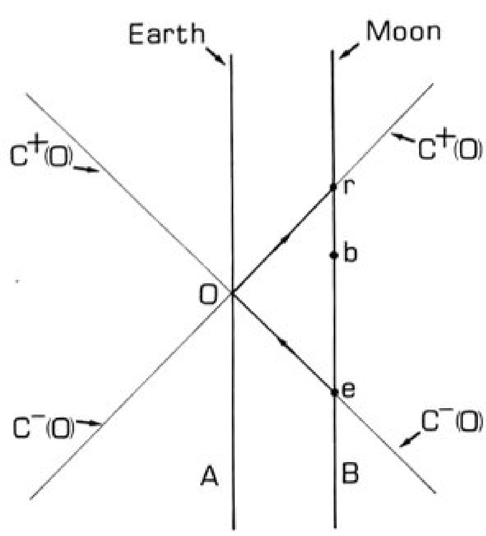


Fig. 1.21 The past and future light cones of an event O in the history of an observer A on the Earth, who (at the event O) sees event e (a threatening boulder starting to roll down) in the history of an astronaut B on the Moon. Observer A immediately sends a warning signal to B; but this arrives at event r, after the boulder has just hit the astronaut at the event b in his history. Because b is outside the future light cone of O, the observer at O cannot influence what happens there.

the surface of the Earth, the resulting delays in communication are negligible. They become significant either when large distances or times are involved, or if the time-scales involved in some process are such that the speed of light is a significant limiting factor. One example is supercomputers: an ultimate limit is imposed on their possible speed of calculation because information cannot be conveyed from one part of the computer to another at speeds greater than the speed of light; this limits the number of calculations that can be performed per second. For this reason, distances between their components must be kept small; thus supercomputers of the future will be small machines.

Exercises

1.6 A satellite takes survey pictures of a square region of the Earth, 800 km in width, from 300 km. above the Earth's surface. What is the delay from the centre of the image to the edge? (Regard the Earth's surface as flat in order to simplify the calculation).

• 1.7 Suppose that a 'mind reader' in London claims to know what his twin brother in New Zealand says at any moment, within less than one-hundredth of a second after a word is uttered. Is there anything extraordinary about this claim? [The radius of the Earth is about 6000 km.]

• 1.8 A rocket R moves in the z direction relative to an observer A on Mars, at a speed v where $v/c = \frac{1}{2}$; their positions coincide at $t = 0$. Plot the world-lines of A and R in a (t, Z) diagram. The rocket emits light signals in both the forward and backward directions at $t = 2$ sec; draw the corresponding light rays in your space–time diagram. The observer A signals to the rocket at the time $t = 1$ sec; what is the earliest time he can expect to get a reply? [All distances and times are measured in the reference frame of the observer A.]

1.9 Draw a diagram to illustrate the fact that the 'past' (i.e. the past light cone and its interior) of any point P on any world-line, always includes the 'past' of any earlier point Q on that world-line. Interpret this result in physical terms.

Computer Exercise 1

Write a program that will either (a) take as input a spatial distance D (in miles or km) and give as output the time T (in seconds, minutes, or hours) for light to travel that distance; or (b) take as input a light travel time T, and give as output the corresponding distance D. Try the program for suitable distances on the Earth, and in the solar system.

Now alter the program to print out additionally the rescaled distance $D1 = D/c$, where c is the speed of light. Notice the simplification achieved. [This corresponds to use of coordinates X, Y, Z discussed above, for which the speed of light is unity. Your output should always state the units of time and distance being used.]

1.3 Relative motion in special relativity

We have seen that even in Newtonian theory, two observers in relative motion will, in general, have different views of space–time. We have also