Lecture Notes on Special Relativity

prepared by

J D Cresser
Department of Physics
Macquarie University

8th August 2005

Contents

Chapter 1

Introduction: What is Relativity?

UNTIL the end of the 19th century it was believed that Newton's three Laws of Motion and the associated ideas about the properties of space and time provided a basis on which the motion of matter could be completely understood. However, the formulation by Maxwell of a unified theory of electromagnetism disrupted this comfortable state of affairs – the theory was extraordinarily successful, yet at a fundamental level it seemed to be inconsistent with certain aspects of the Newtonian ideas of space and time. Ultimately, a radical modification of these latter concepts, and consequently of Newton's equations themselves, was found to be necessary. It was Albert Einstein who, by combining the experimental results and physical arguments of others with his own unique insights, first formulated the new principles in terms of which space, time, matter and energy were to be understood. These principles, and their consequences constitute the Special Theory of Relativity. Later, Einstein was able to further develop this theory, leading to what is known as the General Theory of Relativity. Amongst other things, this latter theory is essentially a theory of gravitation.

Relativity (both the Special and General theories), quantum mechanics, and thermodynamics are the three major theories on which modern physics is based. What is unique about these three theories, as distinct from say the theory of electromagnetism, is their generality. Embodied in these theories are general principles which all more specialized or more specific theories are required to satisfy. Consequently these theories lead to general conclusions which apply to all physical systems, and hence are of enormous power, as well as of fundamental significance. The role of relativity appears to be that of specifying the properties of space and time, the arena in which all physical processes take place.

It is perhaps a little unfortunate that the word 'relativity' immediately conjures up thoughts about the work of Einstein. The idea that a principle of relativity applies to the properties of the physical world is very old: it certainly predates Newton and seems to have been first stated concisely by Galileo, though some of the ideas were already around at the time of Aristotle (who apparently did not believe in the principle). What the principle of relativity essentially states is the following:

> *The laws of physics take the same mathematical form in all frames of reference moving with constant velocity with respect to one another.*

Explicitly recognized in this statement is the empirical fact that the laws of nature, almost without exception, can be expressed in the form of mathematical equations. Why this should be so is a profound issue that is not fully understood, but it is nevertheless the case that doing so offers the most succinct way of summarizing the observed behaviour of a physical system under reproducible experimental conditions. What the above statement is then saying can be ascertained as follows.

Consider a collection of experimenters, (or, as they are often referred to, observers) each based in laboratories moving at constant velocities with respect to one another, and each undertaking a series of experiments designed to lead to a mathematical statement of a particular physical law, such as the response of a body to the application of a force. According to the principle of relativity, the final form of the equations derived (in this case, Newton's laws) will be found to have exactly the same form for all experimenters.

It should be understood that whilst the mathematical form of the laws will be the same, the actually data obtained by each experimenter – even if they are monitoring the same physical event – will not necessarily be numerically the same. For instance, the point in space where two bodies collide, and the time at which this collision occurs, will not necessarily be assigned the same coordinates by all experimenters. However, there is invariably a mathematical relationship between such data obtained by the different observers. In the case of Newtonian relativity these transformation equations constitute the so-called Galilean transformation. Using these transformation equations, the mathematical statement of any physical law according to one observer can be translated into the law as written down by another observer. The principle of relativity then requires that the transformed equations have exactly the same form in all frames of reference moving with constant velocity with respect to one another, in other words that the physical laws are the same in all such frames of reference.

This statement contains concepts such as 'mathematical form' and 'frame of reference' and 'Galilean transformation' which we have not developed, so perhaps it is best at this stage to illustrate its content by a couple of examples. In doing so it is best to make use of an equivalent statement of the principle, that is:

> *Given two observers A and B moving at a constant velocity with respect to one another, it is not possible by any experiment whatsoever to determine which of the observers is 'at rest' or which is 'in motion'.*

First consider an example from 'everyday experience' – a train carriage moving smoothly at a constant speed on a straight and level track – this is a 'frame of reference', an idea that will be better defined later. Suppose that in a carriage of this train there is a pool table and suppose you were a passenger on this carriage, and you decided to play a game of pool. One of the first things that you would notice is that in playing any shot, you would have to make no allowance whatsoever for the motion of the train. Any judgement of how to play a shot as learned by playing the game at home, or in the local pool hall, would apply equally well on the train, irrespective of how fast the train was moving. If we consider that what is taking place here is the innate application of Newton's Laws to describe the motion and collision of the pool balls, we see that no adjustment has to be made to these laws when playing the game on the moving train.

This argument can be turned around. Suppose the train windows are covered, and the carriage is well insulated so that there is no vibration or noise – i.e. there is no immediate evidence to the senses as to whether or not the train is in motion. It might nevertheless still be possible to determine if the train is in motion by carrying out an experiment, such as playing a game of pool. But, as described above, a game of pool proceeds in exactly the same way as if it were being played back home – no change in shot-making is required. There is no indication from this experiment as to whether or not the train is in motion. There is no way of knowing whether, on pulling back the curtains, you are likely to see the countryside hurtling by, or to find the train sitting at a station. In other words, by means of this experiment which, in this case, involves Newton's Laws of motion, it is not possible to determine whether or not the train carriage is moving, an outcome entirely consistent with the principle of relativity.

This idea can be extended to encompass other laws of physics. To this end, imagine a collection of spaceships with engines shut off, all drifting through space. Each space ship constitutes a 'frame of reference'. On each of these ships a series of experiments is performed: a measurement of the half life of uranium 235, a measurement of the outcome of the collision of two billiard balls, an experiment in thermodynamics, e.g. a measurement of the boiling point of water under normal atmospheric pressure, a measurement of the speed of light radiating from a nearby star: any conceivable experiment. If the results of these experiments are later compared, what is found is that in all cases (within experimental error) the results are identical. For instance, we do not find that on one space ship water boils at 100°C, on another hurtling towards the first it boils at 150°C and on another hurtling away from the first, it boils at 70°C. In other words, the various laws of physics being tested here yield exactly the same results for all the spaceships, in accordance with the principle of relativity.

Thus, quite generally, the principle of relativity means that it is not possible, by considering any physical process whatsoever, to determine whether or not one or the other of the spaceships is 'in motion'. The results of all the experiments are the same on all the space ships, so there is nothing that definitely singles out one space ship over any other as being the one that is stationary. It is true that from the point of view of an observer on any one of the space ships that it is the others that are in motion. But the same statement can be made by an observer in *any* space ship. All that we can say for certain is that the space ships are in relative motion, and not claim that one of them is 'truly' stationary, while the others are all 'truly' moving.

This principle of relativity was accepted (in somewhat simpler form i.e. with respect to the mechanical behaviour of bodies) by Newton and his successors, even though Newton postulated that underlying it all was 'absolute space' which defined the state of absolute rest. He introduced the notion in order to cope with the difficulty of specifying with respect to what an accelerated object is being accelerated. To see what is being implied here, imagine space completely empty of all matter except for two masses joined by a spring. Now suppose that the arrangement is rotated around an axis through the centre of the spring, and perpendicular to the spring. As a consequence, the masses will undergo acceleration. Naively, in accordance with our experience, we would expect that the masses would pull apart. But why should they? How do the masses 'know' that they are being rotated? There are no 'signposts' in an otherwise empty universe that would indicate that rotation is taking place. By proposing that there existed an absolute space, Newton was able to claim that the masses are being accelerated with respect to this absolute space, and hence that they would separate in the way expected for masses in circular motion. But this was a supposition made more for the convenience it offered in putting together his Laws of motion, than anything else. It was an assumption that could not be substantiated, as Newton was well aware – he certainly felt misgivings about the concept! Other scientists were more accepting of the idea, however, with Maxwell's theory of electromagnetism for a time seeming to provide some sort of confirmation of the concept.

One of the predictions of Maxwell's theory was that light was an electromagnetic wave that travelled with a speed $c \approx 3 \times 10^8$ ms^{-1}. But relative to what? Maxwell's theory did not specify any particular frame of reference for which light would have this speed. A convenient resolution to this problem was provided by an already existing assumption concerning the way light propagated through space. That light was a form of wave motion was well known – Young's interference experiments had shown this – but the Newtonian world view required that a wave could not propagate through empty space: there must be present a medium of some sort that vibrated as the waves passed, much as a tub of jelly vibrates as a wave travels through it. The proposal was therefore made that space was filled with a substance known as the ether whose purpose was to be the medium that vibrated as the light waves propagated through it. It was but a small step to then propose that this ether was stationary with respect to Newton's absolute space, thereby solving the problem of what the frame of reference was in which light had the speed c. Furthermore, in keeping with the usual ideas of relative motion, the thinking was then that if you were to travel relative to the ether towards a beam of light, you would measure its speed to be greater than c, and less than c if you travelled away from the beam. It then came as an enormous surprise when it was found experimentally that this was not, in fact, the case.

This discovery was made by Michelson and Morley, who fully accepted the ether theory, and who, quite reasonably, thought it would be a nice idea to try to measure how fast the earth was moving through the ether. But the result they found was quite unexpected. Irrespective of the position of the earth in its orbit around the sun, the result was always zero, which made no sense at all: surely somewhere in the orbit the Earth would have to be moving relative to the ether. To put it another way, they measured the speed of light always to be the same value c no matter what the relative motion might be of the Earth with respect to the ether. In our spaceship picture, this is equivalent to all the spaceships obtaining the same value for the speed of light radiated by the nearby star irrespective of their motion relative to the star. This result is completely in conflict with the rule for relative velocities, which in turn is based on the principle of relativity as enunciated by Galileo. Thus the independence of the speed of light on the motion of the observer seems to take on the form of an immutable law of nature, and yet it is apparently inconsistent with the principle of relativity. Something was seriously amiss, and it was Einstein who showed how to get around the problem, and in doing so he was forced to conclude that space and time had properties undreamt of in the Newtonian world picture.

The first contribution made by Einstein was to raise to the level of a postulate the observation that the speed of light was apparently independent of the state of motion of its source, and this, along with the principle of relativity presented above leads to the Special Theory of Relativity. This theory is concerned almost entirely with physical processes as observed from reference frames moving at constant velocities with respect to each other, so-called inertial frames of reference, and incorporates the fact that the results of the hypothetical experiments described above will all be independent of the state of motion of the experimenters. This is an outcome which it is best to understand at a fundamental level in terms of the mathematical forms taken by the laws of nature. All laws of nature appear to have expression in mathematical form, and, as mentioned earlier, the principle of relativity can be understood as saying that the equations describing a law of nature take the same mathematical form in all frames of reference moving at a constant velocity with respect to each other, and moreover, the velocity of the reference frame does not appear anywhere in these equations. But in order to guarantee that the principle of relativity holds true for all physical processes, including the postulate concerning the constancy of the speed of light, Einstein was forced to propose, along with a new perspective on the properties of space and time, modified versions of the familiar Newtonian concepts of force, momentum and energy, leading, amongst other things, to the famous equation $E = mc^2$.

Much later (1915), after a long struggle, Einstein produced a generalization of this theory in which it was required that the laws of physics should be the same in *all* frames of reference whether in constant relative motion, or undergoing acceleration, or even accelerating different amounts in different places. This amounts to saying that any physical process taking place in space and time should proceed in a fashion that takes no account of the reference frame used to describe it. In other words, it ought to be possible to write down the laws of physics in terms of quantities that make no mention whatsoever of any particular reference frame. In accomplishing this task, Einstein was able to show that the force of gravity could be understood as a reflection of underlying geometrical properties of space and time – that space and time can be considered as a single geometric entity that can exhibit curvature.

All these ideas, and a lot more besides, have to be presented in a much more rigorous form. It is this perspective on relativity in terms of the mathematical statements of the laws of physics that is developed here, and an important starting point is pinning down the notion of a frame of reference.

Chapter 2

Frames of Reference

PHYSICAL processes either directly or indirectly involve the dynamics of particles and/or fields moving or propagating through space and time. As a consequence, almost all of the fundamental laws of physics involve position and time in some way or other e.g. Newton's second law of motion

$$\mathbf{F} = m\mathbf{a} \tag{2.1}$$

when applied to a particle responding to the action of a force will yield the position of the particle as a function of time. Likewise, Maxwell's equations will yield the wave equation

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0 \tag{2.2}$$

for the propagation of a light wave through space and time. Implicit in these statements of these fundamental physical laws is the notion that we have at hand some way of measuring or specifying or labelling each point in space and and different instants in time. So, in order to describe in a quantitative fashion the multitude of physical processes that occur in the natural world, one of the important requirements is that we be able to specify where and when events take place in space and time. By 'event' we could be referring to something that occurs at an instant in time at a point in space, or, more colloquially, over a localized interval in time, and in a localized region in space. An evening of opera under the stars with a glass or two of fine wine shared with good company is one example of an event. Such things as a star exploding as a supernova, or a match being struck and flaring up briefly, or two billiard balls colliding with each other, or a space probe passing through one of the rings of Saturn, or a radiactive nucleus emitting a beta particle – all could be understand as being 'events', and in each case we could specify where the event occurred and at what time it took place, provided, of course, we had some means of measuring these quantities: the opera took place at a vineyard 200km to the north from home and just inland from the coast, beginning last Saturday at 8:00pm, or the radioactive decay was of an atom at a certain position in a metallic crystal, with the time of the emission registered by a Geiger counter. Whatever the circumstance, we specify the where and when of an event by measuring its position relative to some conveniently chosen origin, and using a clock synchronized in some agreed fashion with all other clocks, to specify the time. This combination of a means of measuring the position of events, and the time at which they occur, constitutes what is referred to as a frame of reference.

Of course, for the purposes of formulating a mathematical statement of a physical law describing, say, the motion of a particle through space, or the properties of an electromagnetic or some other field propagating through space, a precise way of specifying the where and when of events is required, that is, the notion of a frame of reference, or reference frame, must be more carefully defined.

2.1 Constructing an Arbitrary Reference Frame

A frame of reference can be constructed in essentially any way, provided it meets the requirements that it labels in a unique fashion the position and the time of the occurrence of any event that might occur. A convenient way of imagining how this might be done is to suppose that all of space is filled with a three dimensional lattice or scaffolding – something like a fishing net, perhaps. The idea is illustrated in Fig. 2.1, though in two dimensions only. The net need not be rigid, and the spacing between adjacent points where the coordinate lines cross need not be the same everywhere.

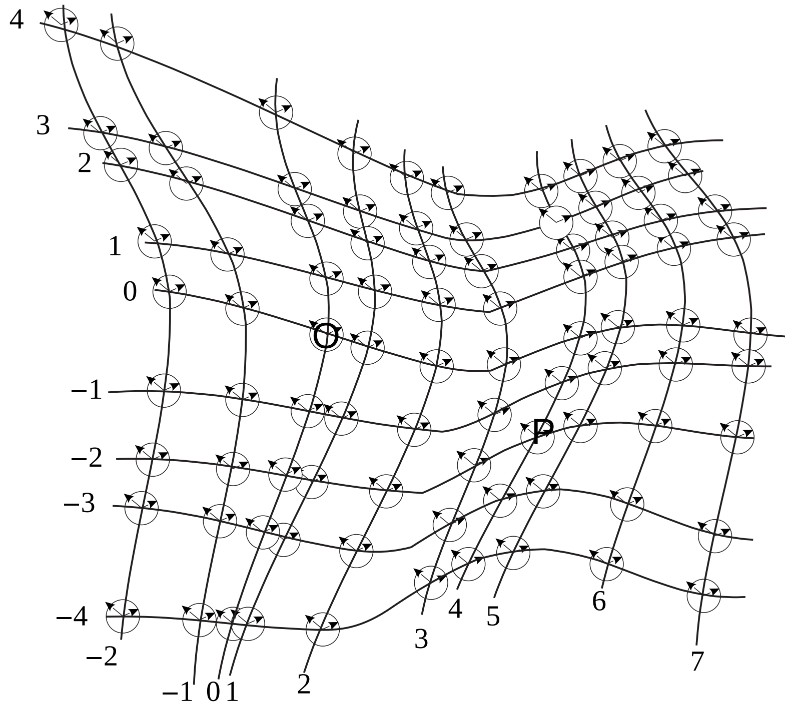

Figure 2.1: A possible network of coordinate lines to specify position in two dimensional space, with clocks attached to each intersection point. Using the labelling attached to this reference system, the origin O is at $(0,0)$ and the point P will have the spatial coordinates $(4,-2)$. The time of an event occurring at P will then be registered by the clock attached to P. Points in space not at the intersection points will be (approximately) labelled by the coordinates of the nearest intersection point, and the time of an event occurring at such a point would be as recorded by the nearest clock. The accuracy with which these positions and times are recorded would be increased by using a finer network of coordinate lines and clocks with better time resolution.

We now give labels to each intersection point on this three dimensional net. We could label each such point in any way we like but it is convenient to do this in some systematic fashion, as illustrated in Fig. 2.1. In the end, of course, we end up labelling each intersection point by a triplet of three real numbers, the usual coordinates of a point in three dimensional space, with one arbitrary point on the net chosen as the origin, and the coordinates of any other crossing point counted off along the netting in some predetermined way.

To complete the picture, we also imagine that attached to the net at each intersection point is a clock. In the same way that we do not necessarily require the net to be rigid, or uniform, we do not necessarily require these clocks to run at the same rate (which is the temporal equivalent of the intersection points on the net not being equally spaced). We do not even require the clocks to be synchronized in any way. At this stage, the role of these clocks is simply to provide us with a specification of the time at which an event occurs in the neighbourhood of the site to which the clock is attached. Thus, if an event occur in space, such as a small supernova flaring up, burn marks will be left on the netting, and the clock closest to the supernova will grind to a halt, so that it reads the time at which the event occurred, while the coordinates of the crossing point closest to where the burn marks appear will give the position of the event.

It does not seem to be a particularly useful state of affairs to have completely arbitrary netting, i.e. it would be far more useful to design the net so that the crossing points are evenly spaced along the threads. This could be done, for instance, by laying out rods of some predetermined length and marking off some convenient separation between the crossing points – and the finer the scale the better. Like wise, it is not particularly useful to have a whole host of clocks ticking away independently of one another[1], particularly if we want to compare whether one event occurring at some point in space occurs earlier or later than some other event occurring elsewhere, and if we want to specify how far apart in time they occur. In other words, it would be preferable to arrange for these clocks to be synchronized in some way. This, however, is not at all a straightforward procedure and in some cases not even possible! It is easy enough to synchronize clocks at the same point in space – the problem is coming up with a way of doing so for clocks at different points in space to be synchronized. In some circumstances it is possible to carry out this synchronization, thereby assigning a global time throughout the reference frame. In such cases, one possible procedure is to suppose that a whole collection of identical clocks are gathered at one point, say the origin of coordinates, and there they are all synchronized to some 'master clock'. These clocks are then carried at an exceedingly slow rate (since, as we will see later, moving clocks 'run slow', and this can affect the synchronization) and distributed around the reference frame – a process known as adiabatic synchronization. Some adjustment may be necessary to the rates of each of these clocks, depending on any gravitational field present, and then we are done [2].

Gravity-free space (i.e. the situation described by special relativity) is one important situation where this synchronization procedure is possible. An expanding isotropic universe is another. There are, however, circumstances in which this cannot be done, such as in space-time around a rotating black hole, or more exotic still, in Gödel's model of a rotating universe.

This combination of clocks and netting thus gives us one possible frame of reference with which to specify the positions and times at which events occur in space and time. With this frame of reference, we could, for instance, plot the position of a particle moving through space as a function of time: just imagine that the particle is highly radioactive so it leaves burn marks on the netting, and stops any closely nearby clock it passes. After the particle has passed by, someone (the observer) clambers along the netting and notes down the coordinates of all the burn marks, and the times registered on the clocks closest to each such mark, takes all this data back to his laboratory, and plots position as a function of time. The result is a depiction of the path of the particle according to this frame of reference.

By using a discrete net and clocks that have a finite time interval between ticks we can only represent the positions and times of occurrence of events to the accuracy determined by how fine the netting is and how long this interval is between clock ticks. But as we believe that space and time are both continuous quantities (though quantum mechanics may have something to say about this), we can suppose that we can get a better approximation to the position and time of

[1] Two chronometers the captain had,
One by Arnold that ran like mad,
One by Kendal in a walnut case,
Poor devoted creature with a hangdog face.

Arnold always hurried with a crazed click-click
Dancing over Greenwich like a lunatic,
Kendal panted faithfully his watch-dog beat,
Climbing out of Yesterday with sticky little feet.

[2] Carrying out the procedure of setting length scales and synchronizing clocks is actually trickier than it seems. For instance, the rods have to be at rest with respect to the net at the location where the the distance is to be marked off, and the rods cannot be too long. In fact, in curved space time, they need to be infinitesimal in length, or at least very short compared to the length scale of the curvature of spacetime in its vicinity. Both the setting of length scales and synchronizing of clocks can be achieved by the use of light signals, but we will not be concerning ourselves with these issues.

occurrence of an event by imagining a finer netting and clocks with shorter intervals ‘between ticks’. Ultimately we would end up with an infinitesimally fine net, and clocks whose ticks occur an infinitesimally short interval apart. But in the end we usually do away with this operationally based picture of nets and clocks and rely on the abstract mathematical notion of a reference frame. But the physical meaning of these mathematically idealized reference frames is nevertheless to be found in the approximate pictures conjured up by using these ideas of a network of clocks attached to a three-dimensional scaffolding filling all of space. When the going gets tough it is often useful to return to the notion of a reference frame defined in this way.

We can set up any number of such reference frames, each with its own coordinate network and set of coordinate clocks. We have (almost) total freedom to set up a reference frame any way we like, including different reference frames being in motion, or even accelerating, with respect to one another, and not necessarily in the same way everywhere. But whichever reference frame we use, we can then conduct experiments whose outcomes are expressed in terms of the associated set of coordinates, and express the various laws of physics in terms of the coordinate systems used. So where does the principle of relativity come into the picture here? What this principle is saying, in its most general form, is that since any physical process taking place in space and time ought to proceed in a fashion that takes no account of the reference frame that we use to describe it. In other words. it ought to be possible to write down the laws of physics in terms of quantites that make no mention whatsoever of any particular reference frame. We can already do this for Newtonian mechanics: Newton’s second law can be written as

$$\mathbf{F} = m\frac{d^2\mathbf{r}}{dt^2}$$

i.e. expressed in a way that makes no mention of a reference frame (though note the appearance of a singled out time variable t – the absolute time of Newton). If we had chosen a particular set of axes, we would have

$$F_x = m\frac{d^2x}{dt^2}$$

and so on where the values of the components of **F** depend on the set of axes chosen. Later we will see how physical laws can be expressed in a ‘frame invariant way’ in the context of special relativity, rather than Newtonian physics. Requiring the relativity principle to be true for arbitrary reference frames, along with a further postulate, the principle of equivalence, which essentially states that an object undergoing free fall in a gravitational field is equivalent to the particle being acted on by no forces at all, then leads to general relativity.

2.1.1 Events

Colloquially, an event is something that occurs at a localized region in space over a localized interval in time, or, in an idealized limit, at a point in space at an instant in time. Thus, the motion of a particle through space could be thought of as a continous series of events, while the collision of two particles would be an isolated event, and so on. However, it is useful to release this term ‘event’ from being associated with something happening. After all, the the coordinate network spread throughout space, and the clocks ticking away the hours will still be labelling points in space, along with ‘the time’ at each point in space, irrespective of whether or not anything actually takes place at a particular locality and at a particular time. The idea then is to use the term ‘event’ simply as another name for a point in space and time, this point specified by the spatial coordinates of the point in space, and the reading of a clock at that point.

An event will have different coordinates in different reference frames. It is then important and useful to be able to relate the coordinates of events in one reference frame to the coordinates of the same event in some other reference frame. In Newtonian physics, this relation is provided by the Galilean transformation equations, and in special relativity by the Lorentz transformation. It is the latter transformation law, and special relativity in particular that we will be concerning ourselves with from now on.

2.2 Inertial Frames of Reference

As we have just seen, a reference frame can be defined in a multitude of ways, but quite obviously it would be preferable to use the simplest possible, which brings to mind the familiar Cartesian set of coordinate axes. Thus, suppose we set up a lattice work of rods as illustrated in Fig. 2.2 in which the rods extend indefinitely in all directions. Of dourse, there will be a third array of rods perpendicular to those in the figure in the Z direction.

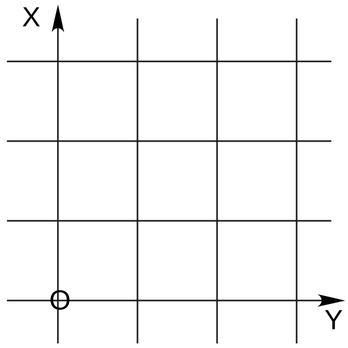

Figure 2.2: Cartesian coordinate system (in two dimensions).

The question then arises: can we in fact do this for *all* of space? From the time of Euclid, and perhaps even earlier, until the 19th century, it was taken for granted that this would be possible, with the rods remaining parallel in each direction out to infinity, even though attempts to prove this from the basic axioms of Euclidean geometry never succeeded. Eventually it was realized by the mathematicians Gauss, Riemann and Lobachevsky that this idea about parallel lines never meeting, while intuitively plausible, was not in fact necesssarily true. It was perfectly possible to construct geometries wherein ‘parallel’ lines could either meet, or diverge, when sufficiently far extended without resulting in any mathematical inconsistencies. Practically, what this means is that by taking short lengths Δl of rod and joining them together in such a way that each length is parallel to the one before, by a process known as parallel transport, then the extended rods could in fact come closer together or become increasingly separated, see Fig. 2.3.

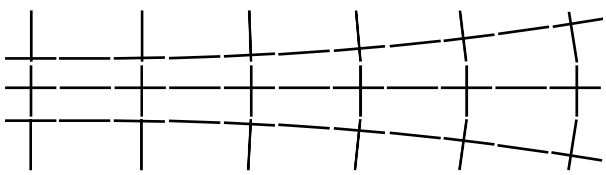

Figure 2.3: Coordinate array constructed by laying down infinitesimal segments of length Δl. Each segment is laid down parallel to its neighbour by a process known as parallel transport. The intrinsic curvature of the underlying space is revealed by the array of rods diverging (as here) or possibly coming closer together.

Such behaviour is indicative of space being intrinsically curved, and there is no apriori reason why space ought to be flat as Euclid assumed, i.e. space could possess some kind of curvature. In fact, Gauss attempted to measure the curvature of space by measuring the area of a very large triangle whose vertices were taken to be the peaks of three widely separated mountain peaks in the Alps. The intention was to see if the area of the triangle came out to be either bigger or smaller than that which would be expected on the basis of Euclidean geometry – either result would have been an indicator of curvature of space. To within the probably very large experimental error of the experiment, no evidence of curvature was found. Riemann, a student of Gauss, surmised that the curvature of space was somehow related to the force of gravity, but he was missing one important ingredient – it is the curvature of space and time together that gives rise to gravity, as Einstein was able to show. Gravitational forces are thus rather peculiar forces in comparison to the other forces of nature such as electromagnetic forces or the nuclear forces in that they are not due to the action of some external influence exerting its effect within pre-existing space and time, but rather is associated with the intrinsic properties of spacetime itself. Thus, we cannot presume that the arrangement of rods as indicated in Fig. 2.2 can be extended indefinitely throughout all of space. However, over a sufficiently localized region of space it can be shown to be the case that such an arrangement is possible, i.e. it is possible to construct frames of reference of a particularly simple form: a lattice of mutually perpendicular rigid rods. In the presence of gravity, it is not possible to extend such a lattice throughout all space, essentially because space is curved. But that is a topic for consideration in general relativity. For gravity free space, these rods can be presumed to extend to infinity in all directions. In addition, it is possible to associate with this lattice a collection of synchronized identical clocks positioned at the intersection of these rods. We will take this simple lattice structure as our starting point for discussing special relativity. That this can be done is a signature of what is known as flat space-time.

First of all we can specify the positions of the particle in space by determining its coordinates relative to a set of mutually perpendicular axes X, Y, Z. In practice this could be done by choosing our origin of coordinates to be some convenient point and imagining that rigid rulers – which we can also imagine to be as long as necessary – are laid out from this origin along these three mutually perpendicular directions. The position of the particle can then be read off from these rulers, thereby giving the three position coordinates (x, y, z) of the particle.

By this means we can specify *where* the particle is. As discussed in Section 2, in order to specify *when* it is at a particular point in space we stretch our imagination further and imagine that in addition to having rulers to measure position, we also have at each point in space a clock, and that these clocks have all been *synchronized* in some way. The idea is that with these clocks we can tell when a particle is at a particular position in space simply by reading off the time indicated by the clock at that position.

According to our 'common sense' notion of time, it would appear sufficient to have only one set of clocks filling all of space. Thus, no matter which set of moving rulers we use to specify the position of a particle, we always use the clocks belonging to this single vast set to tell us when a particle is at a particular position. In other words, there is only one 'time' for all the position measuring set of rulers. This time is the same time independent of how the rulers are moving through space. This is the idea of universal or absolute time due to Newton. However, as Einstein was first to point out, this idea of absolute time is untenable, and that the measurement of time intervals (e.g. the time interval between two events such as two supernovae occurring at different positions in space) will in fact differ for observers in motion relative to each other. In order to prepare ourselves for this possibility, we shall suppose that *for each possible set of rulers* – including those fixed relative to the ground, or those moving with a subatomic particle and so on, there are a *different* set of clocks. Thus the position measuring rulers carry their own set of clocks around with them. The clocks belonging to each set of rulers are of course synchronized with respect to each other. Later on we shall see how this synchronization can be achieved. The idea now is that relative to a particular set of rulers we are able to specify where a particle is, and by looking at the clock (belonging to that set of rulers) at the position of the particle, we can specify when the particle is at that position. Each possible collection of rulers and associated clocks constitutes what is known as a frame of reference or a reference frame.

In many texts reference is often made to an observer in a frame of reference whose job apparently is to make various time and space measurements within this frame of reference. Unfortunately, this conjures up images of a person armed with a stopwatch and a pair of binoculars sitting at the origin of coordinates and peering out into space watching particles (or planets) collide, stars explode and so on. This is not the sense in which the term observer is to be interpreted. It is important to realise that measurements of time are made using clocks which are positioned at the spatial point at which an event occurs. Any centrally positioned observer would have to take account of the time of flight of a signal to his or her observation point in order to calculate the actual time of occurrence of the event. One of the reasons for introducing this imaginary ocean of clocks is to avoid such a complication. Whenever the term observer arises it should be interpreted as meaning the reference frame itself, except in instances in which it is explicitly the case that the observations of an isolated individual are under consideration.

If, as measured by one particular set of rulers and clocks (i.e. frame of reference) a particle is observed to be at a position at a time t (as indicated by the clock at (x, y, z)), we can summarize this information by saying that the particle was observed to be at the point (x, y, z, t) in space-time. The motion of the particle relative to this frame of reference would be reflected in the particle being at different positions (x, y, z) at different times t, see Fig. 2.4.

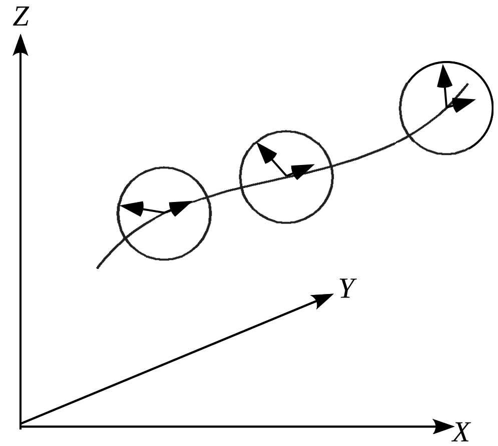

Figure 2.4: Path of a particle as measured in a frame of reference. The clocks indicate the times at which the particle passed the various points along the way.

For instance in the simplest non-trivial case we may find that the particle is moving at constant speed v in the direction of the positive X axis, i.e. $x = vt$. Finally, we could consider the frame of reference whose spatial origin coincides with the particle itself. In this last case, the position of the particle *does not change* since it remains at the spatial origin of its frame of reference. However, the clock associated with this origin keeps on ticking so that the particle's coordinates in space-time are $(0, 0, 0, t)$ with t the time indicated on the clock at the origin, being the only quantity that changes. If a particle remains stationary relative to a particular frame of reference, then that frame of reference is known as the *rest frame* for the particle.

Of course we can use frames of reference to specify the where and when of things other than the position of a particle at a certain time. For instance, the point in space-time at which an explosion occurs, or where and when two particles collide etc., can also be specified by the four numbers (x, y, z, t) relative to a particular frame of reference. In fact any event occurring in space and time can be specified by four such numbers whether it is an explosion, a collision or the passage of a particle through the position (x, y, z) at the time t. For this reason, the four numbers (x, y, z, t) together are often referred to as an *event*.

2.2.1 Newton's First Law of Motion

Having established how we are going to measure the coordinates of a particle in space and time, we can now turn to considering how we can use these ideas to make a statement about the physical properties of space and time. To this end let us suppose that we have somehow placed a particle in the depths of space far removed from all other matter. It is reasonable to suppose that a particle so placed is *acted on by no forces whatsoever*[3]. The question then arises: 'What kind of motion is this

[3]It is not necessary to define what we mean by force at this point. It is sufficient to presume that if the particle is far removed from all other matter, then its behaviour will in no way be influenced by other matter, and will instead be in

particle undergoing?' In order to determine this we have to measure its position as a function of time, and to do this we have to provide a reference frame. We could imagine all sorts of reference frames, for instance one attached to a rocket travelling in some complicated path. Under such circumstances, the path of the particle as measured relative to such a reference frame would be very complex. However, it is at this point that an assertion can be made, namely that for certain frames of reference, the particle will be travelling in a particularly simple fashion – a straight line at constant speed. This is something that has not and possibly could not be confirmed experimentally, but it is nevertheless accepted as a true statement about the properties of the motion of particles in the absences of forces. In other words we can adopt as a law of nature, the following statement:

> *There exist frames of reference relative to which a particle acted on by no forces moves in a straight line at constant speed.*

This essentially a claim that we are making about the properties of spacetime. It is also simply a statement of Newton's First Law of Motion. A frame of reference which has this property is called an inertial frame of reference, or just an inertial frame.

Gravity is a peculiar force in that if a reference frame is freely falling under the effects of gravity, then any particle also freely falling will be observed to be moving in a straight line at constant speed relative to this freely falling frame. Thus freely falling frames constitute inertial frames of reference, at least locally.

---

response to any inherent properties of space (and time) in its vicinity.

Chapter 3

Newtonian Relativity

The arguments in the previous Chapter do not tell us whether there is one or many inertial frames of reference, nor, if there is more than one, does it tell us how we are to relate the coordinates of an event as observed from the point-of-view of one inertial reference frame to the coordinates of the same event as observed in some other. These transformation laws are essential if we are to compare the mathematical statements of the laws of physics in different inertial reference frames. The transformation equations that are derived below are the mathematical basis on which it can be shown that Newton's Laws are consistent with the principle of relativity. In establishing the latter, we can show that there is in fact an infinite number of inertial reference frames.

3.1 The Galilean Transformation

To derive these transformation equations, consider an inertial frame of reference S and a second reference frame S' moving with a velocity v_x relative to S.

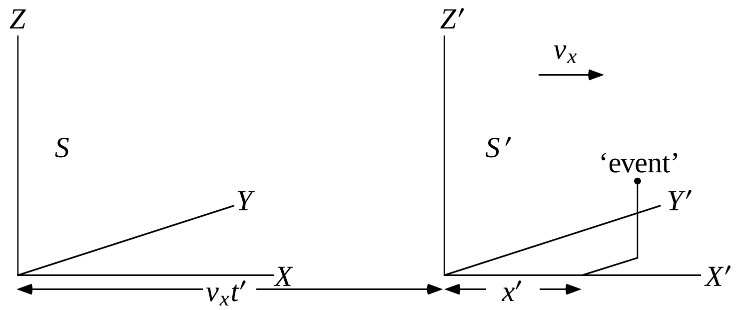

Figure 3.1: A frame of reference S' is moving with a velocity v_x relative to the inertial frame S. An event occurs with spatial coordinates (x, y, z) at time t in S and at (x', y', z') at time t' in S'.

Let us suppose that the clocks in S and S' are set such that when the origins of the two reference frames O and O' coincide, all the clocks in both frames of reference read zero i.e. $t = t' = 0$. According to 'common sense', if the clocks in S and S' are synchronized at $t = t' = 0$, then they will always read the same, i.e. $t = t'$ always. This, once again, is the absolute time concept introduced in Section 2.2. Suppose now that an event of some kind, e.g. an explosion, occurs at a point (x', y', z', t') according to S'. Then, by examining Fig. 3.1, according to S, it occurs at the point

$$x = x' + v_x t', \quad y = y', \quad z = z'$$

and at the time

$$t = t' \tag{3.1}$$

These equations together are known as the Galilean Transformation, and they tell us how the coordinates of an event in one inertial frame S are related to the coordinates of the same event as measured in another frame S' moving with a constant velocity relative to S.

Now suppose that in inertial frame S, a particle is acted on by no forces and hence is moving along the straight line path given by:

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{u}t \tag{3.2}$$

where $\mathbf{u}$ is the velocity of the particle as measured in S. Then in S', a frame of reference moving with a velocity $\mathbf{v} = v_x\mathbf{i}$ relative to S, the particle will be following a path

$$\mathbf{r}' = \mathbf{r}_0 + (\mathbf{u} - \mathbf{v})t' \tag{3.3}$$

where we have simply substituted for the components of $\mathbf{r}$ using Eq. (3.1) above. This last result also obviously represents the particle moving in a straight line path at constant speed. And since the particle is being acted on by no forces, S' is also an inertial frame, and since $\mathbf{v}$ is arbitrary, there is in general an infinite number of such frames.

Incidentally, if we take the derivative of Eq. (3.3) with respect to t, and use the fact that $t = t'$, we obtain

$$\mathbf{u}' = \mathbf{u} - \mathbf{v} \tag{3.4}$$

which is the familiar addition law for relative velocities.

It is a good exercise to see how the inverse transformation can be obtained from the above equations. We can do this in two ways. One way is simply to solve these equations so as to express the primed variables in terms of the unprimed variables. An alternate method, one that is mpre revealing of the underlying symmetry of space, is to note that if S' is moving with a velocity v_x with respect to S, then S will be moving with a velocity $-v_x$ with respect to S' so the inverse transformation should be obtainable by simply exchanging the primed and unprimed variables, and replacing v_x by $-v_x$. Either way, the result obtained is

$$\left.\begin{aligned} x' &= x - v_x t \\ y' &= y \\ z' &= z \\ t' &= t. \end{aligned}\right\} \tag{3.5}$$

3.2 Newtonian Force and Momentum

Having proposed the existence of a special class of reference frames, the inertial frames of reference, and the Galilean transformation that relates the coordinates of events in such frames, we can now proceed further and study whether or not Newton's remaining laws of motion are indeed consistent with the principle of relativity. First we need a statement of these two further laws of motion.

3.2.1 Newton's Second Law of Motion

It is clearly the case that particles do not always move in straight lines at constant speeds relative to an inertial frame. In other words, a particle can undergo acceleration. This deviation from uniform motion by the particle is attributed to the action of a force. If the particle is measured in the inertial frame to undergo an acceleration $\mathbf{a}$, then this acceleration is a consequence of the action of a force $\mathbf{F}$ where

$$\mathbf{F} = m\mathbf{a} \tag{3.6}$$

and where the mass m is a constant characteristic of the particle and is assumed, in Newtonian dynamics, to be the same in all inertial frames of reference. This is, of course, a statement of Newton's Second Law. This equation relates the force, mass and acceleration of a body as measured relative to a particular inertial frame of reference.

As we indicated in the previous section, there are in fact an infinite number of inertial frames of reference and it is of considerable importance to understand what happens to Newton's Second Law if we measure the force, mass and acceleration of a particle from different inertial frames of reference. In order to do this, we must make use of the Galilean transformation to relate the coordinates (x, y, z, t) of a particle in one inertial frame S say to its coordinates (x', y', z', t') in some other inertial frame S'. But before we do this, we also need to look at Newton's Third Law of Motion.

3.2.2 Newton's Third Law of Motion

Newton's Third Law, namely that to every action there is an equal and opposite reaction, can also be shown to take the same form in all inertial reference frames. This is not done directly as the statement of the Law just given is not the most useful way that it can be presented. A more useful (and in fact far deeper result) follows if we combine the Second and Third Laws, leading to the law of conservation of momentum which is

In the absence of any external forces, the total momentum of a system is constant.

It is then a simple task to show that if the momentum is conserved in one inertial frame of reference, then via the Galilean transformation, it is conserved in all inertial frames of reference.

3.3 Newtonian Relativity

By means of the Galilean Transformation, we can obtain an important result of Newtonian mechanics which carries over in a much more general form to special relativity. We shall illustrate the idea by means of an example involving two particles connected by a spring. If the X coordinates of the two particles are x_1 and x_2 relative to some reference frame S then from Newton's Second Law the equation of motion of the particle at x_1 is

$$m_1 \frac{d^2 x_1}{dt^2} = -k(x_1 - x_2 - l) \tag{3.7}$$

where k is the spring constant, l the natural length of the spring, and m_1 the mass of the particle. If we now consider the same pair of masses from the point of view of another frame of reference S' moving with a velocity v_x relative to S, then

$$x_1 = x_1' + v_x t' \quad \text{and} \quad x_2 = x_2' + v_x t' \tag{3.8}$$

so that

$$\frac{d^2 x_1}{dt^2} = \frac{d^2 x_1'}{dt'^2} \tag{3.9}$$

and

$$x_2 - x_1 = x'_2 - x'_1. \tag{3.10}$$

Thus, substituting the last two results into Eq. (3.7) gives

$$m_1 \frac{d^2 x'_1}{dt'^2} = -k(x'_1 - x'_2 - l) \tag{3.11}$$

Now according to Newtonian mechanics, the mass of the particle is the same in both frames i.e.

$$m_1 = m'_1 \tag{3.12}$$

where m'_1 is the mass of the particle as measured in S'. Hence

$$m'_1 \frac{d^2 x'_1}{dt'^2} = -k(x'_1 - x'_2 - l) \tag{3.13}$$

which is exactly the same equation as obtained in S, Eq. (3.7) except that the variables x_1 and x_2 are replaced by x'_1 and x'_2. In other words, the *form* of the equation of motion derived from Newton's Second Law is the same in both frames of reference. This result can be proved in a more general way than for than just masses on springs, and we are lead to conclude that the mathematical form of the equations of motion obtained from Newton's Second Law are the same in all inertial frames of reference.

Continuing with this example, we can also show that momentum is conserved in all inertial reference frames. Thus, in reference frame S, the total momentum is

$$m_1\dot{x}_1 + m_2\dot{x}_2 = P = \text{ constant.} \tag{3.14}$$

Using Eq. (3.8) above we then see that in S' the total momentum is

$$P' = m'_1\dot{x}'_1 + m'_2\dot{x}'_2 = m_1\dot{x}_1 + m_2\dot{x}_2 - (m_1 + m_2)v_x = P - (m_1 + m_2)v_x \tag{3.15}$$

which is also a constant (but not the same constant as in S – it is not required to be the same constant!!). The analogous result to this in special relativity plays a very central role in setting up the description of the dynamics of a system.

The general conclusion we can draw from all this is that:

Newton's Laws of motion are identical in all inertial frames of reference.

This is the Newtonian (or Galilean) principle of relativity, and was essentially accepted by all physicists, at least until the time when Maxwell put together his famous set of equations. One consequence of this conclusion is that it is not possible to determine whether or not a frame of reference is in a state of motion by any experiment involving Newton's Laws. At no stage do the Laws depend on the velocity of a frame of reference relative to anything else, even though Newton had postulated the existence of some kind of ”absolute space” i.e. a frame of reference which defined the state of absolute rest, and with respect to which the motion of anything could be measured. The existence of such a reference frame was taken for granted by most physicists, and for a while it was thought to be have been uncovered following on from the appearance on the scene of Maxwell's theory of electromagnetism.

3.4 Maxwell's Equations and the Ether

The Newtonian principle of relativity had a successful career till the advent of Maxwell's work in which he formulated a mathematical theory of electromagnetism which, amongst other things, provided a successful physical theory of light. Not unexpectedly, it was anticipated that the equations Maxwell derived should also obey the above Newtonian principle of relativity in the sense that Maxwell's equations should also be the same in all inertial frames of reference. Unfortunately, it was found that this was not the case. Maxwell's equations were found to assume completely different forms in different inertial frames of reference. It was as if $\mathbf{F} = m\mathbf{a}$ worked in one frame of reference, but in another, the law had to be replaced by some bizarre equation like $\mathbf{F}' = m(\mathbf{a}')^2\mathbf{a}'$! In other words it appeared as if Maxwell's equations took a particularly simple form in one special frame of reference, but a quite complicated form in another moving relative to this special reference frame. For instance, the wave equation for light assumed the simple form

$$\frac{\partial^2 E}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0 \tag{3.16}$$

in this 'special frame' S, which is the equation for waves moving at the speed c. Under the Galilean transformation, this equation becomes

$$\frac{\partial^2 E'}{\partial x'^2} - \frac{1}{c^2}\frac{\partial^2 E'}{\partial t'^2} - \frac{2v_x}{c^2}\frac{\partial^2 E'}{\partial x'\partial t'} - \frac{v_x}{c^2}\frac{\partial}{\partial x'}\left[v_x\frac{\partial E'}{\partial x'}\right] = 0 \tag{3.17}$$

for a frame S' moving with velocity v_x relative to S. This 'special frame' S was assumed to be the one that defined the state of absolute rest as postulated by Newton, and that stationary relative to it was a most unusual entity, the ether. The ether was a substance that was supposedly the medium in which light waves were transmitted in a way something like the way in which air carries sound waves. Consequently it was believed that the behaviour of light, in particular its velocity, as measured from a frame of reference moving relative to the ether would be different from its behaviour as measured from a frame of reference stationary with respect to the ether. Since the earth is following a roughly circular orbit around the sun, then it follows that a frame of reference attached to the earth must at some stage in its orbit be moving relative to the ether, and hence a change in the velocity of light should be observable at some time during the year. From this, it should be possible to determine the velocity of the earth relative to the ether. An attempt was made to measure this velocity. This was the famous experiment of Michelson and Morley. Simply stated, they argued that if light is moving with a velocity c through the ether, and the Earth was at some stage in its orbit moving with a velocity v relative to the ether, then light should be observed to be travelling with a velocity $c' = c - v$ relative to the Earth. We can see this by simply solving the wave equation in S:

$$E(x, t) = E(x - ct) \tag{3.18}$$

where we are supposing that the wave is travelling in the positive X direction. If we suppose the Earth is also travelling in this direction with a speed v_x relative to the ether, and we now apply the Galilean Transformation to this expression, we get, for the field $E'(x', t')$ as measured in S', the result

$$E'(x', t') = E(x, t) = E(x' + v_x t' - ct') = E(x' - (c - v_x)t') \tag{3.19}$$

i.e. the wave is moving with a speed $c - v_x$ which is just the Galilean Law for the addition of velocities given in Eq. (3.4).

Needless to say, on performing their experiment – which was extremely accurate – they found that the speed of light was always the same. Obviously something was seriously wrong. Their experiments seemed to say that the earth was not moving relative to the ether, which was manifestly wrong since the earth was moving in a circular path around the sun, so at some stage it had to be moving relative to the ether. Many attempts were made to patch things up while still retaining the same Newtonian ideas of space and time. Amongst other things, it was suggested that the earth dragged the ether in its immediate vicinity along with it. It was also proposed that objects contracted in length along the direction parallel to the direction of motion of the object relative to the ether. This suggestion, due to Fitzgerald and elaborated on by Lorentz and hence known as the Lorentz-Fitzgerald contraction, 'explained' the negative results of the Michelson-Morley experiment, but faltered in part because no physical mechanism could be discerned that would be responsible for the contraction. The Lorentz-Fitzgerald contraction was to resurface with a new interpretation following from the work of Einstein. Thus some momentary successes were achieved, but eventually all these attempts were found to be unsatisfactory in various ways. It was Einstein who pointed the way out of the impasse, a way out that required a massive revision of our concepts of space, and more particularly, of time.

Chapter 4

Einsteinian Relativity

THE difficulties with the Newtonian relativity was overcome by Einstein who made two postulates that lead to a complete restructuring of our ideas of space time, and the dyanamical proerties of matter.

4.1 Einstein's Postulates

The difficulty that had to be resolved amounted to choosing amongst three alternatives:

1. The Galilean transformation was correct and something was wrong with Maxwell's equations.
2. The Galilean transformation applied to Newtonian mechanics only.
3. The Galilean transformation, and the Newtonian principle of relativity based on this transformation were wrong and that there existed a new relativity principle valid for both mechanics and electromagnetism that was not based on the Galilean transformation.

The first possibility was thrown out as Maxwell's equations proved to be totally successful in application. The second was unacceptable as it seemed something as fundamental as the transformation between inertial frames could not be restricted to but one set of natural phenomena i.e. it seemed preferable to believe that physics was a unified subject. The third was all that was left, so Einstein set about trying to uncover a new principle of relativity. His investigations led him to make two postulates:

1. All the laws of physics are the same in every inertial frame of reference. This postulate implies that there is no experiment whether based on the laws of mechanics or the laws of electromagnetism from which it is possible to determine whether or not a frame of reference is in a state of uniform motion.
2. The speed of light is independent of the motion of its source.

Einstein was inspired to make these postulates through his study of the properties of Maxwell's equations and not by the negative results of the Michelson-Morley experiment, of which he was apparently only vaguely aware. It is this postulate that forces us to reconsider what we understand by space and time.

One immediate consequence of these two postulates is that the speed of light is the same in all inertial frames of reference. We can see this by considering a source of light and two frames of reference, the first frame of reference S' stationary relative to the source of light and the other, S, moving relative to the source of light.

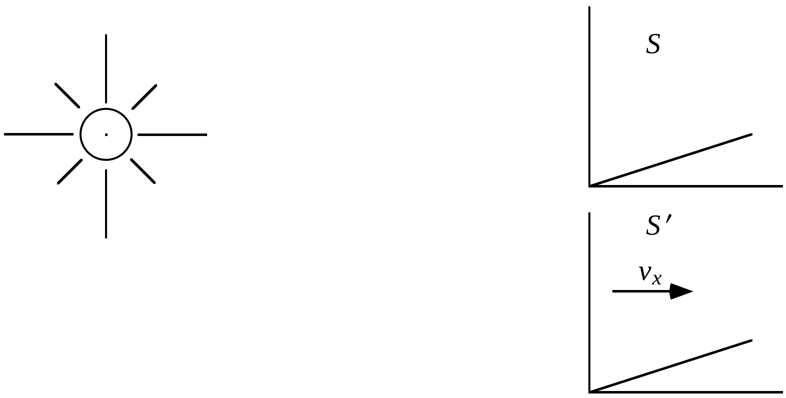

Figure 4.1: A source of light observed from two inertial frames S and S' where S' is moving with a velocity v_x with respect to S.

By postulate 2, S measures the speed of light to be c. However, from postulate 1, this situation is indistinguishable from that depicted in Fig. 4.2

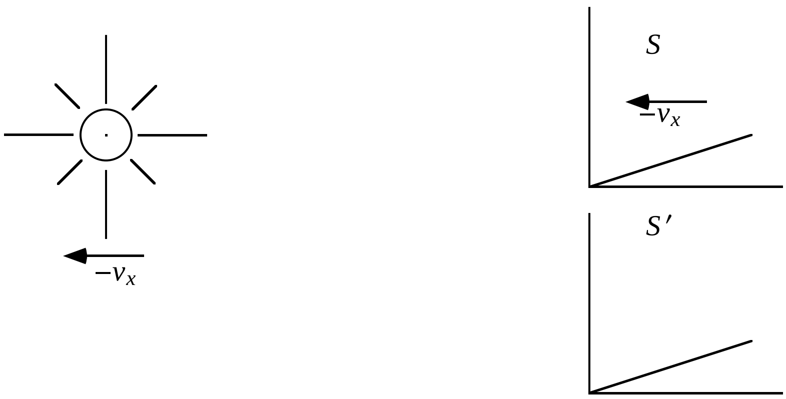

Figure 4.2: The same situation as in Fig. (4.1) except from the point of view of S'.

and by postulate 2, S' must also measure the speed of light to be c. In other words, both reference frames S and measure the speed of light to be c.

Before proceeding further with the consequences of these two rather innocent looking postulates, we have to be more precise about how we go about measuring time in an inertial frame of reference.

4.2 Clock Synchronization in an Inertial Frame

Recall from Section 2.2 that in order to measure the time at which an event occurred at a point in space, we assumed that all of space was filled with clocks, one for each point in space. Moreover, there were a separate set of clocks for each set of rulers so that a frame of reference was defined both by these rulers and by the set of clocks which were carried along by the rulers. It was also stated that all the clocks in each frame of reference were synchronized in some way, left unspecified. At this juncture it is necessary to be somewhat more precise about how this synchronization is to be achieved. The necessity for doing this lies in the fact that we have to be very clear about what we are doing when we are comparing the times of occurrence of events, particularly when the events occur at two spatially separate points.

The procedure that can be followed to achieve the synchronization of the clocks in one frame of reference is quite straightforward. We make use of the fact that the speed of light is precisely known, and is assumed to be always a constant everywhere in free space no matter how it is generated or in which direction it propagates through space. The synchronization is then achieved in the following way. Imagine that at the spatial origin of the frame of reference we have a master clock, and that at some instant $t_0 = 0$ indicated by this clock a spherical flash of light is emitted from the source.

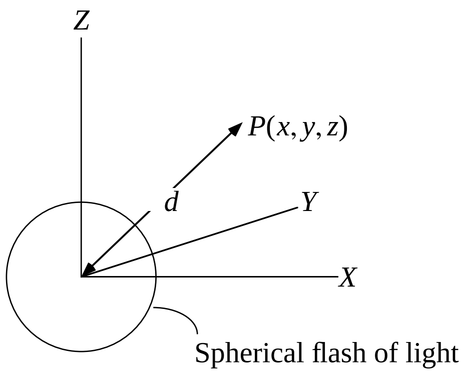

Figure 4.3: A spherical flash of light emitted at $t = 0$ propagates out from the origin, reaching the point P after a time d/c. The clock at P is then set to read $t = d/c$.

The flash of light will eventually reach the point $P(x, y, z)$ situated a distance d from the origin O. When this flash reaches P, the clock at that position is adjusted to read $t = d/c$. And since $d^2 = x^2 + y^2 + z^2$, this means that

$$x^2 + y^2 + z^2 = c^2t^2 \tag{4.1}$$

a result made use of later in the derivation of the Lorentz equations.

This procedure is followed for all the clocks throughout the frame of reference. By this means, the clocks can be synchronized. A similar procedure applies for every frame of reference with its associated clocks.

It should be pointed out that it is not necessary to use light to do this. We could have used any collection of objects whose speed we know with great precision. However it is a reasonable choice to use light since all evidence indicates that light always travels with the same speed c everywhere in space. Moreover, when it comes to comparing observations made in different frames of reference, we can exploit the fact the speed of light always has the same value through postulate 2 above. We do not know as yet what happens for any other objects. In fact, as a consequence of Einstein's second postulate we find that whereas the clocks in one reference frame have all been synchronized to everyone's satisfaction in that frame of reference, it turns out that they are not synchronized with respect to another frame of reference moving with respect to the first. The meaning and significance of this lack of synchronization will be discussed later.

We are now in a position to begin to investigate how the coordinates of an event as measured in one frame of reference are related to the coordinates of the event in another frame of reference. This relationship between the two sets of coordinates constitutes the so-called Lorentz transformation.

4.3 Lorentz Transformation

In deriving this transformation, we will eventually make use of the constancy of the speed of light, but first we will derive the general form that the transformation law must take purely from kinematic/symmetry considerations. Doing so is based on two further assumptions which seem to be entirely reasonable:

Homogeneity: The intrinsic properties of empty space are the same everywhere and for all time. In other words, the properties of the rulers and clocks do not depend on their positions in (empty) space, nor do they vary over time.

Spatial Isotropy: The intrinsic properties of space is the same in all directions. In other words, the properties of the rulers and clocks do not depend on their orientations in empty space.

There is a third, much more subtle condition:

No Memory: The extrinsic properties of the rulers and clocks may be functions of their current states of motion, but not of their states of motion at any other time.

This is not referring to what might happen to a ruler or a clock as a consequence of what it might have done in the past such as, for instance, having undergone such severe acceleration that its inner workings were wrecked. To see what it refers to, we can imagine that we prepare two identical clocks and send one off on an elaborate journey through space and time while the other stays behind. When brought back together, the clocks might not read the same time, but what this postulate is saying is that they will be ticking at the same rate. Similarly for a pair of rulers: they will have the same length when brought back together. Thus we do not have to consider the past history of any of our clocks and rulers when comparing lengths or intervals of time: space and time does not leave a lingering imprint on the objects that live in space and time.

The starting point is to consider two inertial frames S and S' where S' is moving with a velocity v_x relative to S.

Let us suppose that when the two origins coincide, the times on the clocks in each frame of reference are set to read zero, that is $t = t' = 0$. Now consider an event that occurs at the point (x, y, z, t) as measured in S. The same event occurs at (x', y', z', t') in S'. What we are after is a set of equations that relate these two sets of coordinates.

We are going to assume a number of things about the form of these equations, all of which can be fully justified, but which we will introduce more or less on the basis that they seem intuitively reasonable.

First, because the relative motion of the two reference frames is in the X direction, it is reasonable to expect that all distances measured at right angles to the X direction will be the same in both S and S', i.e.[1]

$$y = y' \text{ and } z = z'. \tag{4.2}$$

We now assume that (x, t) and (x', t') are related by the linear transformations

$$x' = Ax + Bt \tag{4.3}$$

$$t' = Cx + Dt. \tag{4.4}$$

Why linear? Assuming that space and time is homogeneous tells us that a linear relation is the only possibility[2]. What it amounts to saying is that it should not matter where in space we choose

[1]If we assumed, for instance, that $z = kz'$, then it would also have to be true that $z' = kz$ if we reverse the roles of S and S', which tells us that $k^2 = 1$ and hence that $k = \pm 1$. We cannot have $z = -z'$ as the coordinate axes are clearly not 'inverted', so we must have $z = z'$.

[2]In general, x' will be a function of x and t, i.e. $x' = f(x, t)$ so that we would have $dx' = f_x dx + f_t dt$ where f_x is the partial derivatve of f with respect to x, and similarly for f_t. Homogeniety then means that these partial derivatives are constants. In other words, a small change in x and t produces the *same* change in x' no matter where in space or time the change takes place.

our origin of the spatial coordinates to be, not should it matter when we choose the orgin of time, i.e. the time that we choose to set as $t = 0$.

Now consider the origin O' of S'. This point is at $x' = 0$ which, if substituted into Eq. (4.3) gives

$$Ax + Bt = 0 \tag{4.5}$$

where x and t are the coordinates of O' as measured in S, i.e. at time t the origin O' has the X coordinate x, where x and t are related by $Ax + Bt = 0$. This can be written

$$\frac{x}{t} = -\frac{B}{A} \tag{4.6}$$

but x/t is just the velocity of the origin O' as measured in S. This origin will be moving at the same speed as the whole reference frame, so then we have

$$-\frac{B}{A} = v_x \tag{4.7}$$

which gives $B = -v_x A$ which can be substituted into Eq. (4.3) to give

$$x' = A(x - v_x t). \tag{4.8}$$

If we now solve Eq. (4.3) and Eq. (4.4) for x and t we get

$$x = \frac{Dx' + v_x At'}{AD - BC} \tag{4.9}$$

$$t = \frac{At' - Cx'}{AD - BC}. \tag{4.10}$$

If we now consider the origin O of the reference frame S, that is, the point $x = 0$, and apply the same argument as just used above, and noting that O will be moving with a velocity $-v_x$ with respect to S', we get

$$-\frac{v_x A}{D} = -v_x \tag{4.11}$$

which then gives

$$A = D \tag{4.12}$$

and hence the transformations Eq. (4.9) and Eq. (4.10) from S' to S will be, after substituting for D and B:

$$\left.\begin{aligned} x &= \frac{(x' + v_x t')}{A + v_x C} \\ t &= \frac{(t' - (C/A)x')}{A + v_x C} \end{aligned}\right\} \tag{4.13}$$

which we can compare with the original transformation from S to S'

$$\left.\begin{aligned} x' &= A(x - v_x t) \\ t' &= A(t + (C/A)x). \end{aligned}\right\} \tag{4.14}$$

At this point we will introduce a notation closer to the conventional notation i.e. we will now write

$$A = \gamma \qquad \text{and} \qquad C/A = K. \tag{4.15}$$

so that the sets of equations above become

$$\left.\begin{aligned} x &= \frac{(x' + v_x t')}{\gamma(1 + v_x K)} \\ t &= \frac{(t' - K x')}{\gamma(1 + v_x K)} \end{aligned}\right\} \tag{4.16}$$

and

$$\left.\begin{aligned} x' &= \gamma(x - v_x t) \\ t' &= \gamma(t + K x). \end{aligned}\right\} \tag{4.17}$$

We now want to make use of some of the symmetry properties listed above to learn more about γ and K. In doing this, it should be understood that the quantities γ and K are not constants. While it is true that they do not depend on x or t, they still potentially depend on v_x. However, the assumed isotropy of space means that γ cannot depend on the *sign* of v_x. If we write $\gamma \equiv \gamma(v_x)$ and $\gamma' \equiv \gamma(-v_x)$, (with a similar meaning for K and K'), this means that[3]

$$\gamma = \gamma'. \tag{4.18}$$

A symmetry property we have already used is that if S' is moving with a velocity v_x relative to S, then S must be moving with velocity $-v_x$ relative to S'. We now make use of this fact to reverse the transformation equations Eq. (4.17) to express x and t in terms of x' and t'. We do this by making the substitutions $v_x \to -v_x$, $x \leftrightarrow x'$, and $t \leftrightarrow t'$, which leads to

$$\left.\begin{aligned} x &= \gamma(x' + v_x t') \\ t &= \gamma(t' + K' x'). \end{aligned}\right\} \tag{4.19}$$

By comparison with Eq. (4.16) we have

$$\gamma = \frac{1}{\gamma(1 + v_x K)} \quad \text{and} \quad \frac{-K}{\gamma(1 + v_x K)} = \gamma K' \tag{4.20}$$

which tells us that

$$\gamma^2 = \frac{1}{1 + v_x K} \quad \text{and} \quad K = -K'. \tag{4.21}$$

The second of these two equations tells us that we can write K as

$$K = -v_x/V^2 \tag{4.22}$$

where V^2 will not depend on the sign of v_x though it could still depend on v_x. We are motivated to write K in this way because by doing so the quantity V will have the units of velocity, which will prove to be convenient later. There is nothing physical implied by doing this, it is merely a mathematical convenience. Thus we have

$$\gamma = \frac{1}{\sqrt{1 - (v_x/V)^2}}. \tag{4.23}$$

[3]To see this, suppose we have a third reference frame S'' which is moving with a velocity $-v_x$ relative to S. We then have the two transformation equations $x' = \gamma(x - v_x t)$ and $x'' = \gamma'(x + v_x t)$. Now suppose some event occurs at the origin of S, i.e. at $x = 0$ at a time t as measured in S. The position of this event as measured in S' will be $x' = -\gamma v_x t$ while, as measured in S'', would be at $x'' = \gamma' v_x t$. By the assumed isotropy of space we ought to have $\gamma v_x t = \gamma' v_x t$ i.e. $\gamma = \gamma'$.

The transformation laws now take the form

$$\left.\begin{aligned} x' &= \frac{x - v_x t}{\sqrt{1-(v_x/V)^2}} \\[2ex] t' &= \frac{t - (v_x/V^2)x}{\sqrt{1-(v_x/V)^2}}. \end{aligned}\right\} \tag{4.24}$$

To determine the dependence of V on v_x, we will suppose there is a further reference frame S'' moving with a velocity $\overline{v}_x$ relative to S'. The same argument as used above can be applied once again to give

$$\left.\begin{aligned} x'' &= \frac{x' - \overline{v}_x t'}{\sqrt{1-(\overline{v}_x/\overline{V})^2}} \\[2ex] t'' &= \frac{t' - (\overline{v}/\overline{V}_x^2)x'}{\sqrt{1-(\overline{v}_x/\overline{V})^2}} \end{aligned}\right\} \tag{4.25}$$

where we have introduced a new parameter $\overline{V}$. If we now substitute for x', y', z' and t' in terms of x, y, z, and t from Eq. (4.24) and rearrange the terms we get

$$x'' = \frac{1 + \overline{v}_x v_x/V^2}{\sqrt{[1 - v_x^2/V^2][1 - \overline{v}_x^2/\overline{V}^2]}}\left[x - \frac{v_x + \overline{v}_x}{1 + v_x\overline{v}_x/V^2}t\right] \tag{4.26}$$

$$t'' = \frac{1 + \overline{v}_x v_x/\overline{V}^2}{\sqrt{[1 - v_x^2/V^2][1 - \overline{v}_x^2/\overline{V}^2]}}\left[t - \frac{v_x/V^2 + \overline{v}_x/\overline{V}^2}{1 + \overline{v}_x v_x/V^2}x\right]. \tag{4.27}$$

This is now the transformation law relating the coordinates of events in S to their coordinates in S''. The transformation equations for y and z contain no surprises and are not included here – it is the transformation equations for x and t that contain the information required. As it stands, these equations, Eq. (4.26) and Eq. (4.27), contain a complicated mess of terms which look very little like the transformations Eq. (4.24) or Eq. (4.25). But, if we are to accept that the transformation between S and S'' should be of the same mathematical form as that between S and S', and that between S' and S'', then we need to look for a condition under which this is true. The easiest way to see what is required is to note that the transformation equations for x and t in Eq. (4.24), are multiplied by the same factor $(1-(v_x/V)^2)^{1/2}$ on the right hand side. So at the very least, we should require that the corresponding factors in Eq. (4.26) and Eq. (4.27) should be equal, putting to one side for the present that fact that even if they are set to equal each other, they still do not look much like $(1-(v_x/V)^2)^{1/2}$! So, proceeding on this basis, we are requiring

$$\frac{1 + \overline{v}_x v_x/V^2}{\sqrt{[1 - v_x^2/V^2][1 - \overline{v}_x^2/\overline{V}^2]}} = \frac{1 + \overline{v}_x v_x/\overline{V}^2}{\sqrt{[1 - v_x^2/V^2][1 - \overline{v}_x^2/\overline{V}^2]}} \tag{4.28}$$

from which it immediately follows that

$$V^2 = \overline{V}^2 \tag{4.29}$$

i.e. the velocity parameter has to be the same for both transformations. As these transformations are arbitrary, we conclude that V^2 has to be a universal constant independent of the relativity velocities of the reference frames. If we then use this fact to simplify all the other terms in Eq. (4.26) and Eq. (4.27) we get

$$\left.\begin{aligned} x'' &= \frac{x - v''_x t}{\sqrt{1-(v''_x/V)^2}} \\ t'' &= \frac{t - v''_x/V^2 x}{\sqrt{1-(v''_x/V)^2}}. \end{aligned}\right\} \tag{4.30}$$

which is exactly of the form of Eq. (4.24) with

$$v''_x = \frac{v_x + \bar{v}_x}{1 + v_x\bar{v}_x/V^2} \tag{4.31}$$

identified as the velocity of the reference frame S'' relative to S.

Thus, the result Eq. (4.24) is the required transformation, with V now shown to be a constant, though one whose value is yet to be determined. This is a remarkable and very general result that depends purely on the assumed homogeneity and isotropy of space. At no stage have we mentioned light, or any other physical quantity for that matter, and yet we have been able to pin down the transformation laws relating coordinate systems for two different inertial frames of reference at least as far as there being only one undetermined quantity left, namely V. This result is one that could have been derived well before Einstein, though the physical or experimental motivation to look for something like this was simply not present.

This parameter V must be looked on as representing some fundamental property of space and time – in fact, it is possible to show from what we have done so far that it represents a 'speed limit' for any moving body that is built into the structure of space and time. More information is needed to determine its value, but we basically have two choices: either V is finite but non-zero, or it is infinite. If we were to choose $V = \infty$, then we find that these transformation equations reduce to the Galilean transformation Eq. (3.1)! However, we have yet to make use of Einstein's second proposal. In doing so we are able to determine V, and find that V has an experimentally determinable, finite value.

To this end, let us suppose that when the two origins coincide, the clocks at O and O' both read zero, and also suppose that at that instant, a flash of light is emitted from the coincident points O and O'. In the frame of reference S this flash of light will be measured as lying on a spherical shell centred on O whose radius is growing at the speed c. However, by the second postulate, in the frame of reference S', the flash of light will also be measured as lying on a spherical shell centred on O' whose radius is also growing at the speed c. Thus, in S, if the spherical shell passes a point P with spatial coordinates (x, y, z) at time t, then by our definition of synchronization we must have:

$$x^2 + y^2 + z^2 = c^2t^2$$

i.e.

$$x^2 + y^2 + z^2 - c^2t^2 = 0. \tag{4.32}$$

The flash of light passing the point P in space at time t then defines an event with spacetime coordinates (x, y, z, t). This event will have a different set of coordinates (x', y', z', t') relative to the frame of reference S' but by our definition of synchronization these coordinates must also satisfy:

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = 0. \tag{4.33}$$

We want to find how the two sets of coordinates (x, y, z, t) and (x', y', z', t') are related in order for both Eq. (4.32) and Eq. (4.33) to hold true. But we know quite generally that these coordinates must be related by the transformation laws Eq. (4.24) obtained above. If we substitute these expressions into Eq. (4.33) we get

$$\left[1-(cv_x/V^2)^2\right]x^2+\left[1-(v_x/V)^2\right]y^2+\left[1-(v_x/V)^2\right]z^2$$
$$-\left[1-(v_x/c)^2\right](ct)^2-2v_x\left[1-(c/V)^2\right]xt=0. \quad (4.34)$$

This equation must reduce to Eq. (4.32). Either by working through the algebra, or simply by trial and error, it is straightforward to confirm that this requires $V = c$, i.e. the general transformation Eq. (4.24) with $V = c$, guarantees that the two spheres of light are expanding at the same rate, that is at the speed c, in both inertial frames of reference. Now writing the quantity γ as

$$\gamma=\frac{1}{\sqrt{1-(v_x/c)^2}} \quad (4.35)$$

we are left with the final form of the transformation law consistent with light always being observed to be travelling at the speed c in all reference frames:

$$\left.\begin{aligned} x' &= \gamma(x-v_xt)\\ y' &= y\\ z' &= z\\ t' &= \gamma(t-(v_x/c^2)x). \end{aligned}\right\} \quad (4.36)$$

These are the equations of the Lorentz transformation. We can find the inverse transformation either by solving Eq. (4.36) for x, y, z, and t in terms of x', y', z', and t', or else by simply recognizing, as was mentioned above in the derivation of this transformation, that if S' is moving with velocity v_x relative to S, then S is moving with velocity $-v_x$ relative to S'. Consequently, all that is required is to exchange the primed and unprimed variables and change the sign of v_x in Eq. (4.36). The result by either method is

$$\left.\begin{aligned} x &= \gamma(x'+v_xt')\\ y &= y'\\ z &= z'\\ t &= \gamma(t'+(v_x/c^2)x'). \end{aligned}\right\} \quad (4.37)$$

These equations were first obtained by Lorentz who was looking for a mathematical transformation that left Maxwell's equations unchanged in form. However he did not assign any physical significance to his results. It was Einstein who first realized the true meaning of these equations, and consequently, with this greater insight, was able to derive them without reference at all to Maxwell's equations. The importance of his insight goes to the heart of relativity. Although the use of a flash of light played a crucial role in deriving the transformation equations, it was introduced as a means by which the value of the unknown parameter V could be determined. The final result simply establishes a connection between the two sets of space-time coordinates associated with a given event, this event being the passage of a flash of light past the point (x, y, z) at time t, as measured in S, or (x', y', z') at time t', as measured in S'. The transformation equations therefore represent a property that space and time must have in order to guarantee that light will always be observed to have the same speed c in all inertial frames of reference. But given that these transformation equations represent an intrinsic property of space and time, it can only be expected that the behaviour of other material objects, which may have nothing whatsoever to do with light, will also be influenced by this fundamental property of space and time. This is the insight that Einstein had, that the Lorentz transformation was saying something about the properties of space and time, and the consequent behaviour that matter and forces must have in order to be consistent with these properties.

Later we will see that the speed of light acts as an upper limit to how fast any material object can travel, be it light or electrons or rocket ships. In addition, we shall see that anything that travels at this speed c will always be observed to do so from all frames of reference. Light just happens to be one of the things in the universe that travels at this particular speed. Subatomic particles called neutrinos also apparently travel at the speed of light, so we could have formulated our arguments above on the basis of an expanding sphere of neutrinos! The constant c therefore represents a characteristic property of space and time, and only less significantly is it the speed at which light travels.

Two immediate conclusions can be drawn from the Lorentz Transformation. Firstly, suppose that , $v_x > c$ i.e. that S' is moving relative to S at a speed greater than the speed of light. In that case we find that $\gamma^2 < 0$ i.e. γ is imaginary so that both position and time in Eq. (4.36) become imaginary. However position and time are both physical quantities which must be measured as real numbers. In other words, the Lorentz transformation becomes physically meaningless if $v_x > c$. This immediately suggests that it is a physical impossibility for a material object to attain a speed greater than c relative to any reference frame S. The frame of reference in which such an object would be stationary will then also be moving at the speed v_x, but as we have just seen, in this situation the transformation law breaks down. We shall see later how the laws of dynamics are modified in special relativity, one of the consequences of this modification being that no material object can be accelerated to a speed greater than c[4].

Secondly, we can consider the form of the Lorentz Transformation in the mathematical limit $v_x << c$. We find that $\gamma \approx 1$ so that Eq. (4.36) becomes the equations of the Galilean Transformation, Eq. (3.1). (Though this also requires that the x dependent term in the time transformation equation to be negligible, which it will be over small enough distances). Thus, at low enough speeds, any unusual results due to the Lorentz transformation would be unobservable.

4.4 Relativistic Kinematics

The Lorentz transformation leads to a number of important consequences for our understanding of the motion of objects in space and time without concern for how the matter was set into motion, i.e. the kinematics of matter. Later, we will look at the consequnces for our understanding of the laws of motion themselves, that is relativistic dynamics.

Perhaps the most startling aspect of the Lorentz Transformation is the appearance of a transformation for time. The result obtained earlier for the Galilean Transformation agrees with, indeed it was based on, our ‘common sense’ notion that time is absolute i.e. that time passes in a manner completely independent of the state of motion of any observer. This is certainly not the case with the Lorentz Transformation which leads, as we shall see, to the conclusion that moving clocks run slow. This effect, called time dilation, and its companion effect, length contraction will now be discussed.

[4]In principle there is nothing wrong with having an object that is initially travelling with a speed greater than c. In this case, c acts as a lower speed limit. Particles with this property, called tachyons, have be postulated to exist, but they give rise to problems involving causality (i.e. cause and effect) which make their existence doubtful.

4.4.1 Length Contraction

The first of the interesting consequences of the Lorentz Transformation is that length no longer has an absolute meaning: the length of an object depends on its motion relative to the frame of reference in which its length is being measured. Let us consider a rod moving with a velocity v_x relative to a frame of reference S, and lying along the X axis. This rod is then *stationary* relative to a frame of reference S' which is also moving with a velocity v_x relative to S.

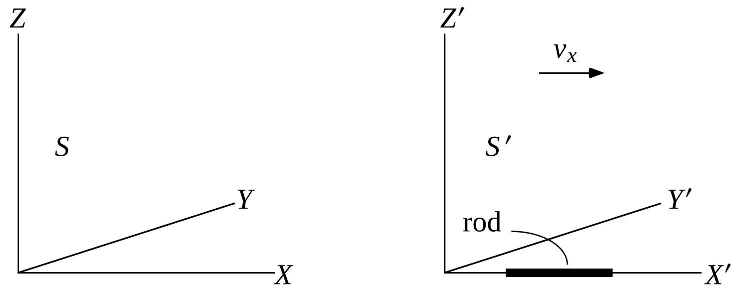

Figure 4.4: A rod of length at rest in reference frame S' which is moving with a velocity v_x with respect to another frame S.

As the rod is stationary in S', the ends of the rod will have coordinates x'_1 and x'_2 which remain fixed as functions of the time in S'. The length of the rod, as measured in S' is then

$$l_0 = x'_2 - x'_1 \tag{4.38}$$

where l_0 is known as the proper length of the rod i.e. l_0 is its length as measured in a frame of reference in which the rod is stationary. Now suppose that we want to measure the length of the rod as measured with respect to S. In order to do this, we measure the X coordinates of the two ends of the rod at the same time t, as measured by the clocks in S. Let x_2 and x_1 be the X coordinates of the two ends of the rod as measured in S at this time t. It is probably useful to be aware that we could rephrase the preceding statement in terms of the imaginary synchronized clocks introduced in Section 2.2 and Section 4.2 by saying that 'the two clocks positioned at x_2 and x_1 both read t when the two ends of the rod coincided with the points x_2 and x_1.' Turning now to the Lorentz Transformation equations, we see that we must have

$$\left.\begin{aligned} x'_1 &= \gamma(x_1 - v_x t)) \\ x'_2 &= \gamma(x_2 - v_x t). \end{aligned}\right\} \tag{4.39}$$

We then define the length of the rod as measured in the frame of reference S to be

$$l = x_2 - x_1 \tag{4.40}$$

where the important point to be re-emphasized is that this length is defined in terms of the positions of the ends of the rods as measured at the same time t in S. Using Eq. (4.39) and Eq. (4.40) we find

$$l_0 = x'_2 - x'_1 = \gamma(x_2 - x_1) = \gamma l \tag{4.41}$$

which gives for l

$$l = \gamma^{-1} l_0 = \sqrt{1 - (v_x/c)^2}\, l_0. \tag{4.42}$$

But for $v_x < c$

$$\sqrt{1 - (v_x/c)^2} < 1 \tag{4.43}$$

so that

$$l < l_0. \tag{4.44}$$

Thus the length of the rod as measured in the frame of reference S with respect to which the rod is moving is shorter than the length as measured from a frame of reference S' relative to which the rod is stationary. A rod will be observed to have its maximum length when it is stationary in a frame of reference. The length so-measured, l_0 is known as its *proper length*.

This phenomenon is known as the Lorentz-Fitzgerald contraction. It is not the consequence of some force 'squeezing' the rod, but it is a real physical phenomenon with observable physical effects. Note however that someone who actually looks at this rod as it passes by will not see a shorter rod. If the time that is required for the light from each point on the rod to reach the observer's eye is taken into account, the overall effect is that of making the rod appear as if it is rotated in space.

4.4.2 Time Dilation

Perhaps the most unexpected consequence of the Lorentz transformation is the way in which our 'commonsense' concept of time has to be drastically modified. Consider a clock C' placed at rest in a frame of reference S' at some point x' on the X axis. Suppose once again that this frame is moving with a velocity v_x relative to some other frame of reference S. At a time t'_1 registered by clock C' there will be a clock C_1 in the S frame of reference passing the position of C':

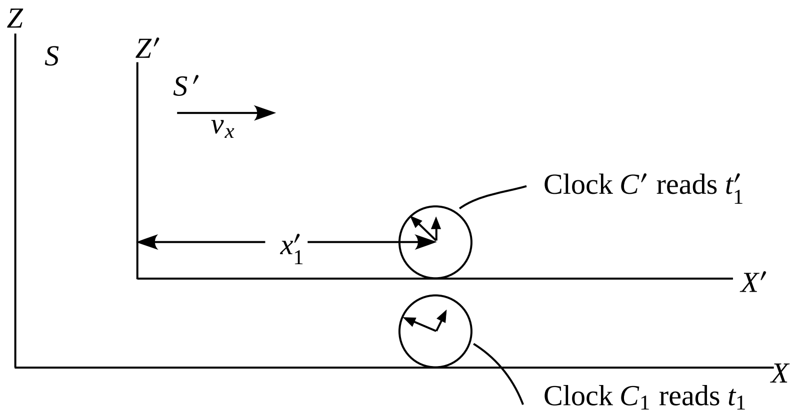

Figure 4.5: Clock C' stationary in S' reads t'_1 when it passes clock C_1 stationary in S, at which instant it reads t_1.

The time registered by C_1 will then be given by the Lorentz Transformation as

$$t_1 = \gamma(t'_1 + v_x x'/c^2). \tag{4.45}$$

Some time later, clock C' will read the time t'_2 at which instant a *different* clock C_2 in S will pass the position x'_1 in S'.

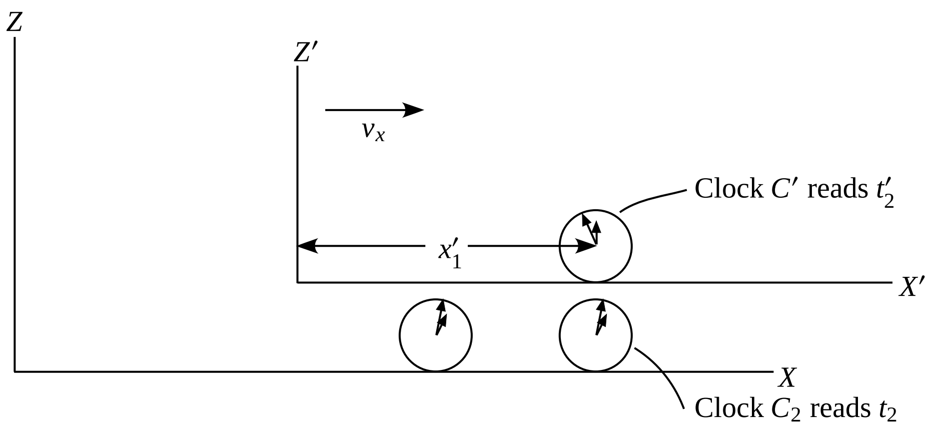

Figure 4.6: Clock C' stationary in S' reads t'_2 when it passes clock C_2 stationary in S, at which instant C_2 reads t_2.

This clock C_2 will read

$$t_2 = \gamma(t_2' + v_x x'/c^2). \tag{4.46}$$

Thus, from Eq. (4.45) and Eq. (4.46) we have

$$\Delta t = t_2 - t_1 = \gamma(t_2' - t_1') = \gamma\Delta t'. \tag{4.47}$$

Once again, since

$$\gamma = \frac{1}{\sqrt{1-(v_x/c)^2}} > 1 \text{ if } v_x < c \tag{4.48}$$

we have

$$\Delta t > \Delta t'. \tag{4.49}$$

In order to interpret this result, suppose that $\Delta t'$ is the time interval between two 'ticks' of the clock C'. Then according to the clocks in S, these two 'ticks' are separated by a time interval Δt which, by Eq. (4.49) is $> \Delta t'$. Thus the time interval between 'ticks' is longer, as measured by the clocks in S, than what it is measured to be in S'. In other words, from the point of view of the frame of reference S, the clock (and all the clocks in S') are running slow. It appears from S that time is passing more slowly in S' than it is in S. This is the phenomenon of *time dilation*. A clock will be observed to run at its fastest when it is stationary in a frame of reference. The clock is then said to be measuring *proper time*.

This phenomenon is just as real as length contraction. One of its best known consequences is that of the increase in the lifetime of a radioactive particle moving at a speed close to that of light. For example, it has been shown that if the lifetime of a species of radioactive particle is measured while stationary in a laboratory to be T', then the lifetime of an identical particle moving relative to the laboratory is found to be given by $T = \gamma T'$, in agreement with Eq. (4.47) above.

Another well known consequence of the time dilation effect is the so-called twin or clock paradox. The essence of the paradox can be seen if we first of all imagine two clocks moving relative to each other which are synchronized when they pass each other. Then, in the frame of reference of one of the clocks, C say, the other clock will be measured as running slow, while in the frame of reference of clock C', the clock C will also be measured to be running slow[5]. This is not a problem until one of the clocks does a U-turn in space (with the help of rocket propulsion, say) and returns to the position of the other clock. What will be found is that the clock that 'came back' will have lost time compared to the other. Why should this be so, as each clock could argue (if clocks could argue) that from its point of view it was the other clock that did the U-turn? The paradox can be resolved in many ways. The essence of the resolution, at least for the version of the clock paradox being considered here, is that there is not complete symmetry between the two clocks. The clock that turns back must have undergone acceleration in order to turn around. The forces associated with this acceleration will only be experienced by this one clock so that even though each clock could argue that it was the other that turned around and came back, it was only one clock that experienced an acceleration. Thus the two clocks have different histories between meetings and it is this asymmetry that leads to the result that the accelerated clock has lost time compared to the other. Of course, we have not shown how the turning around process results in this asymmetry: a detailed analysis is required and will not be considered here.

4.4.3 Simultaneity

Another consequence of the transformation law for time is that events which occur simultaneously in one frame of reference will not in general occur simultaneously in any other frame of reference.

[5]This appears to be paradoxical – how can *both* clocks consider the other as going slow? It should be borne in mind that the clocks C and C' are not being compared directly against one another, rather the time on each clock is being compared against the time registered on the collection of clocks that it passes in the other reference frame

Thus, consider two events 1 and 2 which are simultaneous in S i.e. $t_1 = t_2$, but which occur at two different places x_1 and x_2. Then, in S', the time interval between these two events is

$$\begin{aligned} t_2' - t_1' &= \gamma(t_2 - v_x x_2/c^2) - \gamma(t_1 - v_x x_1/c^2) \\ &= \gamma(x_1 - x_2)v_x/c^2 \\ &\neq 0 \text{ as } x_1 \neq x_2. \end{aligned} \tag{4.50}$$

Here t_1' is the time registered on the clock in S' which coincides with the position x_1 in S at the instant t_1 that the event 1 occurs and similarly for t_2'. Thus events which appear simultaneous in S are not simultaneous in S'. In fact the order in which the two events 1 and 2 are found to occur in will depend on the sign of $x_1 - x_2$ or v_x. It is only when the two events occur at the same point (i.e. $x_1 = x_2$) that the events will occur simultaneously in all frames of reference.

4.4.4 Transformation of Velocities (Addition of Velocities)

Suppose, relative to a frame S, a particle has a velocity

$$\mathbf{u} = u_x\mathbf{i} + u_y\mathbf{j} + u_z\mathbf{k} \tag{4.51}$$

where $u_x = dx/dt$ etc. What we require is the velocity of this particle as measured in the frame of reference S' moving with a velocity v_x relative to S. If the particle has coordinate x at time t in S, then the particle will have coordinate x' at time t' in S' where

$$x = \gamma()x' + v_x t') \text{ and } t = \gamma(t' + v_x x'/c^2). \tag{4.52}$$

If the particle is displaced to a new position $x + dx$ at time $t + dt$ in S, then in S' it will be at the position $x' + dx'$ at time $t' + dt'$ where

$$x + dx = \gamma\,(x' + dx' + v_x(t' + dt'))$$

$$t + dt = \gamma\left(t' + dt' + v_x(x' + dx')/c^2)\right)$$

and hence

$$dx = \gamma(dx' + v_x dt')$$

$$dt = \gamma(dt' + v_x dx'/c^2)$$

so that

$$\begin{aligned} u_x = \frac{dx}{dt} &= \frac{dx' + v_x dt'}{dt' + v_x dx'/c^2} \\ &= \frac{\frac{dx'}{dt'} + v_x}{1 + \frac{v_x}{c^2}\frac{dx'}{dt'}} \\ &= \frac{u_x' + v_x}{1 + v_x u_x'/c^2} \end{aligned} \tag{4.53}$$

where $u_x' = dx'/dt'$ is the X velocity of the particle in the S' frame of reference. Similarly, using $y = y'$ and $z = z'$ we find that

$$u_y = \frac{u_y'}{\gamma(1 + v_x u_x'/c^2)} \tag{4.54}$$

$$u_z = \frac{u_z'}{\gamma(1 + v_x u_x'/c^2)}. \tag{4.55}$$

The inverse transformation follows by replacing $v_x \to -v_x$ interchanging the primed and unprimed variables. The result is

$$\left.\begin{aligned} u'_x &= \frac{u_x - v_x}{1 - v_x u_x/c^2} \\ u'_y &= \frac{u_y}{\gamma(1 - v_x u_x/c^2)} \\ u'_z &= \frac{u_z}{\gamma(1 - v_x u_x/c^2)}. \end{aligned}\right\} \tag{4.56}$$

In particular, if $u_x = c$ and $u_y = u_z = 0$, we find that

$$u'_x = \frac{c - v_x}{1 - v_x/c} = c \tag{4.57}$$

i.e., if the particle has the speed c in S, it has the same speed c in S'. This is just a restatement of the fact that if a particle (or light) has a speed c in one frame of reference, then it has the same speed c in all frames of reference.

Now consider the case in which the particle is moving with a speed that is less that c, i.e. suppose $u_y = u_z = 0$ and $|u_x| < c$. We can rewrite Eq. (4.56) in the form

$$\begin{aligned} u'_x - c &= \frac{u_x - c}{1 - u_x v_x/c^2} - c \\ &= \frac{(c + v_x)(c - v_x)}{c(1 - v_x u_x/c^2)}. \end{aligned} \tag{4.58}$$

Now, if S' is moving relative to S with a speed less than c, i.e. $|v_x| < c$, then along with $|u_x| < c$ it is not difficult to show that the right hand side of Eq. (4.58) is always negative i.e.

$$u'_x - c < 0 \text{ if } |u_x| < c,\ |v_x| < c \tag{4.59}$$

from which follows $u'_x < c$.

Similarly, by writing

$$\begin{aligned} u'_x + c &= \frac{u_x - v_x}{1 - u_x v_x/c^2} + c \\ &= \frac{(c + u_x)(c - v_x)}{c(1 - v_x u_x/c^2)} \end{aligned} \tag{4.60}$$

we find that the right hand side of Eq. (4.60) is always positive provided $|u_x| < c$ and $|v_x| < c$ i.e.

$$u'_x + c > 0 \text{ if } |u_x| < c,\ |v_x| < c \tag{4.61}$$

from which follows $u'_x > -c$. Putting together Eq. (4.59) and Eq. (4.61) we find that

$$|u'_x| < c \text{ if } |u_x| < c \text{ and } |v_x| < c. \tag{4.62}$$

What this result is telling us is that if a particle has a speed less than c in one frame of reference, then its speed is always less than c in any other frame of reference, provided this other frame of reference is moving at a speed less than c. As an example, consider two objects A and B approaching each other, A at a velocity $u_x = 0.99c$ relative to a frame of reference S, and B stationary in a frame of reference S' which is moving with a velocity $v_x = -0.99c$ relative to S.

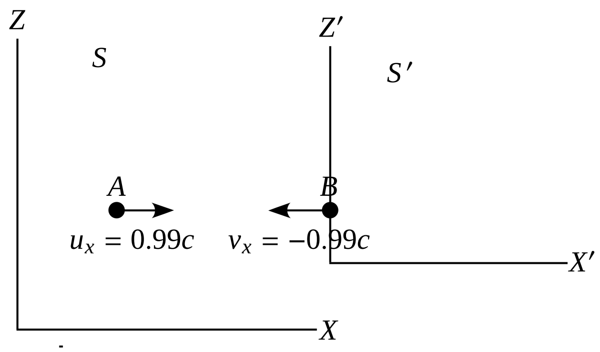

Figure 4.7: Object B stationary in reference frame S' which is moving with a velocity $v_x = -0.99c$ relative to reference frame S. Object A is moving with velocity $u_x = 0.99c$ with respect to reference frame S.

According to classical Newtonian kinematics, B will measure A as approaching at a speed of $1.98c$. However, according to the Einsteinian law of velocity addition, the velocity of A relative to B, i.e. the velocity of A as measured in frame S' is, from Eq. (4.56)

$$u'_x = \frac{0.99c - (-0.99c)}{1 + (0.99)^2} = 0.99995c$$

which is, of course, less than c, in agreement with Eq. (4.62).

In the above, we have made use of the requirement that all speeds be less than or equal to c. To understand physically why this is the case, it is necessary to turn to consideration of relativistic dynamics.

4.5 Relativistic Dynamics

Till now we have only been concerned with kinematics i.e. what we can say about the motion of the particle without consideration of its cause. Now we need to look at the laws that determine the motion i.e. the relativistic form of Newton's Laws of Motion. Firstly, Newton's First Law is accepted in the same form as presented in Section 2.2.1. However two arguments can be presented which indicate that Newton's Second Law may need revision. One argument only suggests that something may be wrong, while the second is of a much more fundamental nature. Firstly, according to Newton's Second Law if we apply a constant force to an object, it will accelerate without bound i.e. up to and then beyond the speed of light. Unfortunately, if we are going to accept the validity of the Lorentz Transformation, then we find that the factor γ becomes imaginary i.e. the factor γ becomes imaginary. Thus real position and time transform into imaginary quantities in the frame of reference of an object moving faster than the speed of light. This suggests that a problem exists, though it does turn out to be possible to build up a mathematical theory of particles moving at speeds greater than c (tachyons).

The second difficulty with Newton's Laws arise from the result, derived from the Second and Third laws, that in an isolated system, the total momentum of all the particles involved is constant, where momentum is defined, for a particle moving with velocity $\mathbf{u}$ and having mass m, by

$$\mathbf{p} = m\mathbf{u} \tag{4.63}$$

The question then is whether or not this law of conservation of momentum satisfies Einstein's first postulate, i.e. with momentum defined in this way, is momentum conserved in all inertial frames of reference? To answer this, we could study the collision of two bodies

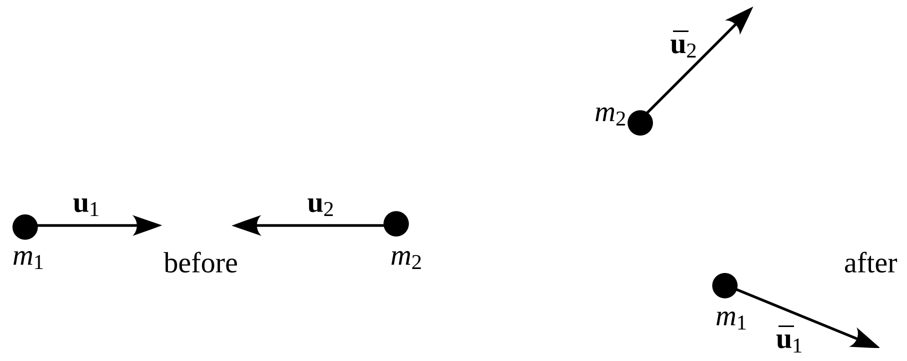

Figure 4.8: Collision between two particles used in discussing the conservation of momentum in different reference frames.

and investigate whether or not we always find that

$$m_1\mathbf{u}_1 + m_2\mathbf{u}_2 = m_1\bar{\mathbf{u}}_1 + m_2\bar{\mathbf{u}}_2 \tag{4.64}$$

in *every inertial frame of reference*. Recall that the velocities must be transformed according to the relativistic laws given by Eq. (4.56). If, however, we retain the Newtonian principle that the mass of a particle is independent of the frame of reference in which it is measured (see Section 3.3) we find that Eq. (4.64) does *not* hold true in all frames of reference – one look at the complex form of the velocity transformation formulae would suggest this conclusion. Thus the Newtonian definition of momentum and the Newtonian law of conservation of momentum are inconsistent with the Lorentz transformation, even though at very low speeds (i.e. very much less than the speed of light) these Newtonian principles are known to yield results in agreement with observation to an exceedingly high degree of accuracy. So, instead of abandoning the momentum concept entirely in the relativistic theory, a more reasonable approach is to search for a generalization of the Newtonian concept of momentum in which the law of conservation of momentum is obeyed in all frames of reference. We do not know beforehand whether such a generalization even exists, and any proposals that we make can only be justified in the long run by the success or otherwise of the generalization in describing what is observed experimentally.

4.5.1 Relativistic Momentum

Any relativistic generalization of Newtonian momentum must satisfy two criteria:

1. Relativistic momentum must be conserved in all frames of reference.
2. Relativistic momentum must reduce to Newtonian momentum at low speeds.

The first criterion must be satisfied in order to satisfy Einstein's first postulate, while the second criterion must be satisfied as it is known that Newton's Laws are correct at sufficiently low speeds. By a number of arguments, the strongest of which being based on arguments concerning the symmetry properties of space and time, a definition for the relativistic momentum of a particle moving with a velocity $\mathbf{u}$ as measured with respect to a frame of reference S, that satisfies these criteria can be shown to take the form

$$\mathbf{p} = \frac{m_0\mathbf{u}}{\sqrt{1 - u^2/c^2}} \tag{4.65}$$

where m_0 is the rest mass of the particle, i.e. the mass of the particle when at rest, and which can be identified with the Newtonian mass of the particle. With this form for the relativistic momentum, Einstein then postulated that, for a system of particles:

The total momentum of a system of particles is always conserved in all frames of reference, whether or not the total number of particles involved is constant.

The above statement of the law of conservation of relativistic momentum generalized to apply to situations in which particles can stick together or break up (that is, be created or annihilated) is only a postulate whose correctness must be tested by experiment. However, it turns out that the postulate above, with relativistic momentum defined as in Eq. (4.65) is amply confirmed experimentally.

We note immediately that, for $u << c$, Eq. (4.65) becomes

$$\mathbf{p} = m_0\mathbf{u} \tag{4.66}$$

which is just the Newtonian form for momentum, as it should be.

It was once the practice to write the relativistic momentum, Eq. (4.65), in the form

$$\mathbf{p} = m\mathbf{u} \tag{4.67}$$

where

$$m = \frac{m_0}{\sqrt{1 - u^2/c^2}} \tag{4.68}$$

which leads us to the idea that the mass of a body (m) increases with its velocity. However, while a convenient interpretation in certain instances, it is not a recommended way of thinking in general since the (velocity dependent) mass defined in this way does not always behave as might be expected. It is better to consider m_0 as being an intrinsic property of the particle (in the same way as its charge would be), and that it is the momentum that is increased by virtue of the factor in the denominator in Eq. (4.65).

Having now defined the relativistic version of momentum, we can now proceed towards setting up the relativistic ideas of force, work, and energy.

4.5.2 Relativistic Force, Work, Kinetic Energy

All these concepts are defined by analogy with their corresponding Newtonian versions. Thus relativistic force is defined as

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \tag{4.69}$$

a definition which reduces to the usual Newtonian form at low velocities. This force will do work on a particle, and the *relativistic work* done by $\mathbf{F}$ during a small displacement $d\mathbf{r}$ is, once again defined by analogy as

$$dW = \mathbf{F} \cdot d\mathbf{r} \tag{4.70}$$

The rate at which $\mathbf{F}$ does work is then

$$P = \mathbf{F} \cdot \mathbf{u} \tag{4.71}$$

and we can introduce the notion of relativistic kinetic energy by viewing the work done by $\mathbf{F}$ as contributing towards the kinetic energy of the particle i.e.

$$P = \frac{dT}{dt} = \mathbf{F} \cdot \mathbf{u} \tag{4.72}$$

where T is the *relativistic kinetic energy* of the particle. We can write this last equation as

$$\frac{dT}{dt} = \mathbf{F}\cdot\mathbf{u} = \mathbf{u}\cdot\frac{d\mathbf{p}}{dt}$$

$$= \mathbf{u}\cdot\frac{d}{dt}\frac{m_0\mathbf{u}}{\sqrt{1-u^2/c^2}}$$

$$= \frac{m_0\mathbf{u}\cdot\dfrac{d\mathbf{u}}{dt}}{\sqrt{1-u^2/c^2}} + \frac{m_0\mathbf{u}\cdot\mathbf{u}u\dfrac{du}{dt}}{c^2\sqrt{1-u^2/c^2}}$$

But

$$\mathbf{u}\cdot\frac{d\mathbf{u}}{dt} = u\frac{du}{dt} \tag{4.73}$$

and hence

$$\frac{dT}{dt} = \left[\frac{m_0}{\sqrt{1-u^2/c^2}} + \frac{m_0u^2/c^2}{\sqrt{(1-u^2/c^2)^3}}\right]u\frac{du}{dt}$$

$$= \frac{m_0}{\sqrt{(1-u^2/c^2)^3}}u\frac{du}{dt}$$

so that we end up with

$$\frac{dT}{dt} = \frac{d}{dt}\left[\frac{m_0c^2}{\sqrt{1-u^2/c^2}}\right]. \tag{4.74}$$

Integrating with respect to t gives

$$T = \frac{m_0c^2}{\sqrt{1-u^2/c^2}} + \text{ constant.} \tag{4.75}$$

By requiring that $T = 0$ for $u = 0$, we find that

$$T = \frac{m_0c^2}{\sqrt{1-u^2/c^2}} - m_0c^2. \tag{4.76}$$

Interestingly enough, if we suppose that $u << c$, we find that, by the binomial approximation[6]

$$\frac{1}{\sqrt{1-u^2/c^2}} = (1-u^2/c^2)^{-\frac{1}{2}} \approx 1 + \frac{u^2}{2c^2} \tag{4.77}$$

so that

$$T \approx m_0c^2(1+u^2/c^2) - m_0c^2 \approx \tfrac{1}{2}m_0c^2 \tag{4.78}$$

which, as should be the case, is the classical Newtonian expression for the kinetic energy of a particle of mass moving with a velocity $\mathbf{u}$.

[6]The binomial approximation is $(1+x)^n \approx 1 + nx$ if $x << 1$.

4.5.3 Total Relativistic Energy

We can now define a quantity E by

$$E = T + m_0c^2 = \frac{m_0c^2}{\sqrt{1-u^2/c^2}}. \tag{4.79}$$

This quantity E is known as the *total relativistic energy* of the particle of rest mass m_0. It is all well and good to define such a thing, but, apart from the neatness of the expression, is there any real need to introduce such a quantity? In order to see the value of defining the total relativistic energy, we need to consider the transformation of momentum between different inertial frames S and S'. To this end consider

$$p_x = \frac{m_0c^2}{\sqrt{1-u^2/c^2}} \tag{4.80}$$

where

$$u = \sqrt{u_x^2 + u_y^2 + u_z^2} \tag{4.81}$$

and where $\mathbf{u}$ is the velocity of the particle relative to the frame of reference S. In terms of the velocity $\mathbf{u}'$ of this particle relative to the frame of reference S' we can write

$$u_x = \frac{u'_x + v_x}{1 + u'_xv_x/c^2} \quad u_y = \frac{u'_y}{\gamma(1 + u'_xv_x/c^2)} \quad u_z = \frac{u'_z}{\gamma(1 + u'_xv_x/c^2)} \tag{4.82}$$

with

$$\gamma = \frac{1}{\sqrt{1 - v_x^2/c^2}} \tag{4.83}$$

as before. After a lot of exceedingly tedious algebra, it is possible to show that

$$\sqrt{1-u^2/c^2} = \frac{\sqrt{1-u'^2/c^2}\,\sqrt{1-v_x^2/c^2}}{1 + u'_xv_x/c^2} \tag{4.84}$$

so that, using Eq. (4.82), Eq. (4.83) and Eq. (4.84) we find

$$p_x = \frac{m_0(u'_x + v_x)}{\sqrt{(1-u'^2/c^2)(1-v_x^2/c^2)}}$$

$$= \gamma\Big[\frac{m_0u'_x}{\sqrt{1-u'^2/c^2}} + v_x\Big(\frac{m_0}{\sqrt{1-u'^2/c^2}}\Big)\Big]$$

which we can readily write as

$$p_x = \gamma[p'_x + v_x(E'/c^2)] \tag{4.85}$$

i.e. we see appearing the total energy E' of the particle as measured in S'.

A similar calculation for p_y and p_z yields

$$p_y = p'_y \text{ and } p_z = p'_z \tag{4.86}$$

while for the energy E we find

$$E = \frac{m_0c^2}{\sqrt{1-u^2/c^2}}$$

$$= \frac{m_0c^2}{\sqrt{1-u'^2/c^2}} \cdot \frac{1 + u'_xv_x/c^2}{\sqrt{1-v_x^2/c^2}}$$

$$= \gamma\Big[\frac{m_0c^2}{\sqrt{1-u'^2/c^2}} + \frac{m_0u'_xv_x}{\sqrt{1-u'^2/c^2}}\Big]$$

which we can write as

$$E = \gamma \left[E' + p'_x v_x\right]. \tag{4.87}$$

Now consider the collision between two particles 1 and 2. Let the X components of momentum of the two particles be p_{1x} and p_{2x} relative to S. Then the total momentum in S is

$$P_x = p_{1x} + p_{2x} \tag{4.88}$$

where P_x is, by conservation of relativistic momentum, a constant, i.e. P_x stays the same before and after any collision between the particles. However

$$p_{1x} + p_{2x} = \gamma \left(p'_{1x} + p'_{2x}\right) + \gamma \left(E'_1 + E'_2\right) v_x/c^2 \tag{4.89}$$

where p'_{1x} and p'_{2x} are the X component of momentum of particles 1 and 2 respectively, while E'_1 and E'_2 are the energies of particles 1 and 2 respectively, all relative to frame of reference S'. Thus we can write

$$P_x = \gamma P'_x + \gamma \left(E'_1 + E'_2\right) v_x/c^2. \tag{4.90}$$

Once again, as momentum is conserved in all inertial frames of reference, we know that P'_x is also a constant i.e. the same before and after any collision. Thus we can conclude from Eq. (4.90) that

$$E'_1 + E'_2 = \text{ constant} \tag{4.91}$$

i.e. the total relativistic energy in S' is conserved. But since S' is an arbitrary frame of reference, we conclude that the total relativistic energy is conserved in all frames of reference (though of course the conserved value would in general be different in different frames of reference). Since, as we shall see later, matter can be created or destroyed, we generalize this to read:

The total relativistic energy of a system of particles is always conserved in all frames of reference, whether or not the total number of particles remains a constant.

Thus we see that conservation of relativistic momentum implies conservation of total relativistic energy in special relativity whereas in Newtonian dynamics, they are independent conditions. Nevertheless, both conditions have to be met in when determining the outcome of any collision between particles, i.e. just as in Newtonian dynamics, the equations representing the conservation of energy and momentum have to be employed.

A useful relationship between energy and momentum can also be established. Its value lies both in treating collision problems and in suggesting the existence of particles with zero rest mass. The starting point is the expression for energy

$$E = \frac{m_0 c^2}{\sqrt{1 - u^2/c^2}} \tag{4.92}$$

from which we find

$$E^2 = \frac{m_0^2 c^4}{1 - u^2/c^2}$$

$$= \frac{m_0 c^4 \left[1 - u^2/c^2 + u^2/c^2\right]}{1 - u^2/c^2}$$

so that

$$E^2 = m_0^2 c^4 + \frac{m_0 u^2}{1 - u^2/c^2} \cdot c^2. \tag{4.93}$$

But

$$\mathbf{p} = \frac{m_0\mathbf{u}}{\sqrt{1 - u^2/c^2}}$$

and hence

$$p^2 = \mathbf{p}\cdot\mathbf{p} = \frac{m_0 u^2}{1 - u^2/c^2}$$

which can be combined with Eq. (4.93) to give

$$E^2 = p^2c^2 + m_0^2c^4. \tag{4.94}$$

We now will use the above concept of relativistic energy to establish the most famous result of special relativity, the equivalence of mass and energy.

4.5.4 Equivalence of Mass and Energy

This represents probably the most important result of special relativity, and gives a deep physical meaning to the concept of the total relativistic energy E. To see the significance of E in this regard, consider the breakup of a body of rest mass m_0 into two pieces of rest masses m_{01} and m_{02}:

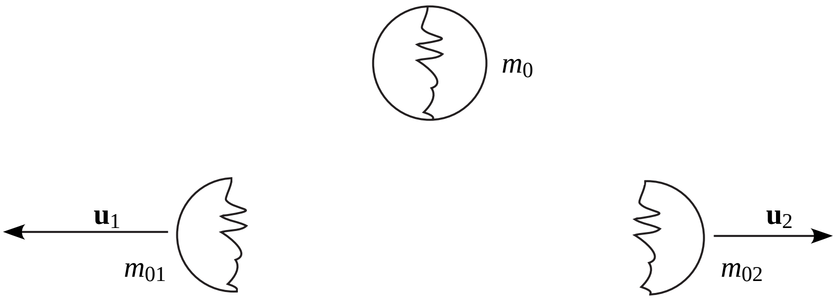

Figure 4.9: Break up of a body of rest mass m_0 into two parts of rest masses m_{01} and m_{02}, moving with velocities $\mathbf{u}_1$ and $\mathbf{u}_2$ relative to the rest frame of the original object.

We could imagine that the original body is a radioactive nucleus, or even simply two masses connected by a coiled spring. If we suppose that the initial body is stationary in some frame S, and the debris flies apart with velocities $\mathbf{u}_1$ and $\mathbf{u}_2$ relative to S then, by the conservation of energy in S:

$$\begin{aligned} E = m_0c^2 &= E_1 + E_2 \\ &= \frac{m_{01}c^2}{\sqrt{1 - u_1^2/c^2}} \frac{m_{02}c^2}{\sqrt{1 - u_2^2/c^2}} \end{aligned}$$

so that

$$\begin{aligned} (m_0 - m_{01} - m_{02})c^2 =& m_{01}c^2\left[\frac{1}{\sqrt{1 - u_1^2/c^2}} - 1\right] \\ &+ m_{02}c^2\left[\frac{1}{\sqrt{1 - u_2^2/c^2}} - 1\right] \\ =& T_1 + T_2 \end{aligned} \tag{4.95}$$

where T_1 and T_2 are the relativistic kinetic energies of the two masses produced. Quite obviously, T_1 and $T_2 > 0$ since

$$\frac{1}{\sqrt{1 - u_1^2/c^2}} - 1 > 0 \tag{4.96}$$

and similarly for the other term and hence

$$m_0 - m_{01} - m_{02} > 0 \tag{4.97}$$

or

$$m_0 < m_{01} + m_{02}. \tag{4.98}$$

What this result means is that the total rest mass of the two separate masses is less than that of the original mass. The difference, Δm say, is given by

$$\Delta m = \frac{T_1 + T_2}{c^2}. \tag{4.99}$$

We see therefore that part of the rest mass of the original body has disappeared, and an amount of kinetic energy given by Δmc^2 has appeared. The inescapable conclusion is that some of the rest mass of the original body has been converted into the kinetic energy of the two masses produced.

The interesting result is that none of the masses involved need to be travelling at speeds close to the speed of light. In fact, Eq. (4.99) can be written, for $u_1, u_2 << c$ as

$$\Delta m = \frac{\frac{1}{2}m_{01}u_1^2 + \frac{1}{2}m_{02}u_2^2}{c^2} \tag{4.100}$$

so that only classical Newtonian kinetic energy appears. Indeed, in order to measure the mass loss Δm, it would be not out of the question to bring the masses to rest in order to determine their rest masses. Nevertheless, the truly remarkable aspect of the above conclusions is that it has its fundamental origin in the fact that there exists a universal maximum possible speed, the speed of light which is built into the structure of space and time, and this structure ultimately exerts an effect on the properties of matter occupying space and time, that is, its mass and energy.

The reverse can also take place i.e. matter can be created out of energy as in, for instance, a collision between particles having some of their energy converted into new particles as in the proton-proton collision

$$p + p \longrightarrow p + p + p + \overline{p}$$

where a further proton and antiproton ($\overline{p}$) have been produced.

A more mundane outcome of the above connection between energy and mass is that rather than talking about the rest mass of a particle, it is often more convenient to talk about its rest energy. A particle of rest mass m_0 will, of course, have a rest energy m_0c^2. Typically the rest energy (or indeed any energy) arising in atomic, nuclear, or elementary physics is given in units of electron volts. One electron volt (eV) is the energy gained by an electron accelerated through a potential difference of 1 volt i.e.

$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ Joules.}$$

An example of the typical magnitudes of the rest energies of elementary particles is that of the proton. With a rest mass of $m_p = 1.67 \times 10^{-33}$ kg, the proton has a rest energy of

$$m_pc^2 = 938.26 \text{ MeV.}$$

4.5.5 Zero Rest Mass Particles

For a single particle, rest mass m_0, its momentum p and energy E are related by the expression:

$$E^2 = p^2c^2 + M_0^2c^4.$$

This result allows us to formally take the limit of $m_0 \to 0$ while keeping E and p fixed. The result is a relationship between energy and momentum for a particle of zero rest mass. In this limit, with $E, p \neq 0$, we have

$$E = pc = |\mathbf{p}|c \tag{4.101}$$

i.e. p is the magnitude of the momentum vector $\mathbf{p}$. If we rearrange Eq. (4.79) to read

$$E\sqrt{1 - u^2/c^2} = m_0c^2$$

and if we then let $m_0 \to 0$ with $E \neq 0$, we must have

$$\sqrt{1 - u^2/c^2} \to 0$$

so that, in the limit of $m_0 \to 0$, we find that

$$u = c. \tag{4.102}$$

Thus, if there exists particles of zero rest mass, we see that their energy and momentum are related by Eq. (4.101) and that they always travel at the speed of light. Particles with zero rest mass need not exist since all that we have presented above is a mathematical argument. However it turns out that they do indeed exist: the photon (a particle of light) and the neutrino, though recent research in solar physics seems to suggest that the neutrino may in fact have a non-zero, but almost immeasurably tiny mass. Quantum mechanics presents us with a relationship between frequency f of a beam of light and the energy of each photon making up the beam:

$$E = hf = \hbar\omega \tag{4.103}$$

Chapter 5

Geometry of Flat Spacetime

THE theory of relativity is a theory of space and time and as such is a geometrical theory, though the geometry of space and time together is quite different from the Euclidean geometry of ordinary 3-dimensional space. Nevertheless it is found that if relativity is recast in the language of vectors and "distances" (or preferably "intervals") a much more coherent picture of the content of the theory emerges. Indeed, relativity is seen to be a theory of the geometry of the single entity, 'spacetime', rather than a theory of space and time. Furthermore, without the geometrical point-of-view it would be next to impossible to extend special relativity to include transformations between arbitrary (non-inertial) frames of reference, which ultimately leads to the general theory of relativity, the theory of gravitation. In order to set the stage for a discussion of the geometrical properties of space and time, a brief look at some of the more familiar ideas of geometry, vectors etc in ordinary three dimensional space is probably useful.

5.1 Geometrical Properties of 3 Dimensional Space

For the present we will not be addressing any specifically relativistic problem, but rather we will concern ourselves with the issue of fixing the position in space of some arbitrary point. To do this we could, if we wanted to, imagine a suitable set of rulers so that the position of a point P can be specified by the three coordinates (x, y, z) with respect to this coordinate system, which we will call R.

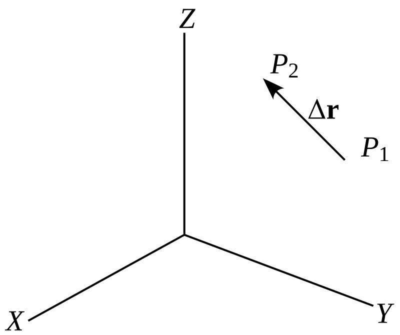

Figure 5.1: A displacement vector $\Delta\mathbf{r}$ in space with an arbitrary coordinate system R.

If we then consider two such points P_1 with coordinates (x_1, y_1, z_1) and P_2 with coordinates (x_2, y_2, z_2) then the line joining these two points defines a vector $\Delta\mathbf{r}$ which we can write in component form with respect to R as

$$\Delta\mathbf{r} \doteq \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix}_R \tag{5.1}$$

where the subscript R is to remind us that the components are specified relative to the set of coordinates R. Why do we need to be so careful? Obviously, it is because we could have, for instance, used a different set of axes R' which have been translated and rotated relative to the first:

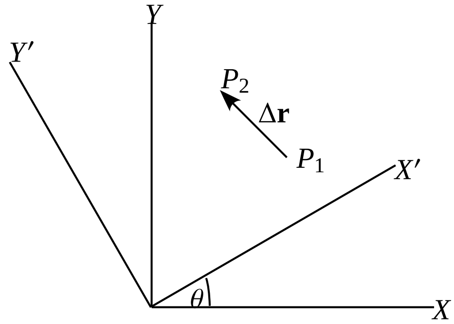

Figure 5.2: Displacement vector and two coordinate systems rotated with respect to each other about Z axis through angle θ. The vector has an existence independent of the choice of coordinate systems.

In this case the vector $\Delta\mathbf{r}$ will have new components, but the vector itself will *still be the same vector* i.e.

$$\Delta\mathbf{r} \doteq \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix}_R \doteq \begin{pmatrix} x'_2 - x'_1 \\ y'_2 - y'_1 \\ z'_2 - z'_1 \end{pmatrix}_{R'} \tag{5.2}$$

or

$$\Delta\mathbf{r} \doteq \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}_R \doteq \begin{pmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{pmatrix}_{R'} \tag{5.3}$$

So the components themselves are meaningless unless we know with respect to what coordinate system they were determined. In fact, the lack of an absolute meaning of the components unless the set of axes used is specified means that the vector $\Delta\mathbf{r}$ is not so much 'equal' to the column vector as 'represented by' the column vectors – hence the use of the dotted equal sign '$\doteq$' to indicate 'represented by'.

The description of the vector in terms of its components relative to some coordinate system is something done for the sake of convenience. Nevertheless, although the components may change as we change coordinate systems, what does not change is the vector itself, i.e. it has an existence independent of the choice of coordinate system. In particular, the length of $\Delta\mathbf{r}$ and the angles between any two vectors $\Delta\mathbf{r}_1$ and $\Delta\mathbf{r}_2$ will be the same in any coordinate system.

While these last two statements may be obvious, it is important for what comes later to see that they also follow by explicitly calculating the length and angle between two vectors using their components in two different coordinate systems. In order to do this we must determine how the coordinates of $\Delta\mathbf{r}$ are related in the two different coordinate systems. We can note that the displacement of the two coordinate systems with respect to each other is immaterial as we are considering differences between vectors thus we only need to worry about the rotation which we have, for simplicity, taken to be through an angle θ about the Z axis (see the above diagram). The transformation between the sets of coordinates can then be shown to be given, in matrix form, by

$$\begin{pmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{pmatrix}_{R'} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}_R \tag{5.4}$$

Using this transformation rule, we can show that

$$(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 \tag{5.5}$$

where each side of this equation is, obviously, the (distance)^2 between the points P_1 and P_2. Further, for any two vectors $\Delta\mathbf{r}_1$ and $\Delta\mathbf{r}_2$ we find that

$$\Delta x_1 \Delta x_2 + \Delta y_1 \Delta y_2 + \Delta z_1 \Delta z_2 = \Delta x'_1 \Delta x'_2 + \Delta y'_1 \Delta y'_2 + \Delta z'_1 \Delta z'_2 \tag{5.6}$$

where each side of the equation is the scalar product of the two vectors i.e. $\Delta\mathbf{r}_1 \cdot \Delta\mathbf{r}_2$. This result tells us that the angle between is the same in both coordinate systems. Thus the transformation Eq. (5.4) is consistent with the fact that the length and relative orientation of these vectors is independent of the choice of coordinate systems, as it should be.

It is at this point that we turn things around and say that *any* quantity that has three components that transform in exactly the same way as $\Delta\mathbf{r}$ under a rotation of coordinate system constitutes a *three-vector*. An example is force, for which

$$\begin{pmatrix} F'_x \\ F'_y \\ F'_z \end{pmatrix}_{R'} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix}_R \tag{5.7}$$

for two coordinate systems R and R' rotated relative to each other by an angle θ about the Z-axis. Other three-vectors are electric and magnetic fields, velocity, acceleration etc. Since the transformation matrix in Eq. (5.7) is identical to that appearing in Eq. (5.4), any three-vector is guaranteed to have the same length (i.e. magnitude) and orientation irrespective of the choice of coordinate system. In other words we can claim that such a three vector has an absolute meaning independent of the choice of coordinate system used to determine its components.

5.2 Space Time Four-Vectors

What we do now is make use of the above considerations to introduce the idea of a vector to describe the separation of two events occurring in spacetime. The essential idea is to show that the coordinates of an event have transformation properties analogous to Eq. (5.4) for ordinary three-vectors, though with some surprising differences. To begin, we will consider two events E_1 and E_2 occurring in spacetime. For event E_1 with coordinates (x_1, y_1, z_1, t_1) in frame of reference S and (x'_1, y'_1, z'_1, t'_1) in S', these coordinates are related by the Lorentz transformation which we will write as

$$\left.\begin{aligned} ct'_1 &= \gamma c t_1 - \frac{\gamma v_x}{c} x_1 \\ x'_1 &= -\frac{\gamma v_x}{c} c t_1 + \gamma x_1 \\ y'_1 &= y_1 \\ z'_1 &= z_1 \end{aligned}\right\} \tag{5.8}$$

and similarly for event E_2. Then we can write

$$\left.\begin{aligned} c\Delta t' &= c(t'_2 - t'_1) &&= \gamma c \Delta t - \frac{\gamma v_x}{c}\Delta x \\ \Delta x' &= x'_2 - x'_1 &&= -\frac{\gamma v_x}{c} c\Delta t + \gamma \Delta x \\ \Delta y' &= \Delta y \\ \Delta z' &= \Delta z \end{aligned}\right\} \tag{5.9}$$

which we can write as

$$\begin{pmatrix} c\Delta t' \\ \Delta x' \\ \Delta y' \\ \Delta z' \end{pmatrix}_{S'} = \begin{pmatrix} \gamma & -\gamma v_x/c & 0 & 0 \\ -\gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}_{S} . \tag{5.10}$$

It is tempting to interpret this equation as relating the components with respect to a coordinate system S' of some sort of 'vector', to the components with respect to some other coordinate system S, of the same vector. We would be justified in doing this if this 'vector' has the properties, analogous to the length and angle between vectors for ordinary three-vectors, which are independent of the choice of reference frame. It turns out that it is 'length' defined as

$$(\Delta s)^2 = (c\Delta t)^2 - \left[(\Delta x)^2 + (\Delta y)^2 + (\Delta y)^2\right] = (c\Delta t)^2 - (\Delta \mathbf{r})^2 \tag{5.11}$$

that is invariant for different reference frames i.e.

$$(\Delta s)^2 = (c\Delta t)^2 - \left[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\right] = (c\Delta t')^2 - \left[(\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2\right] \tag{5.12}$$

This invariant quantity Δs is known as the *interval* between the two events E_1 and E_2. Obviously Δs is analogous to, but fundamentally different from, the length of a three-vector in that it can be positive, zero, or negative. We could also talk about the 'angle' between two such 'vectors' and show that

$$(c\Delta t_1)(c\Delta t_2) - [\Delta x_1 \Delta x_2 + \Delta y_1 \Delta y_2 + \Delta z_1)\Delta z_2] \tag{5.13}$$

has the same value in all reference frames. This is analogous to the scalar product for three-vectors. The quantity defined by

$$\Delta \vec{s} = \begin{pmatrix} c\Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \tag{5.14}$$

is then understood to correspond to a property of spacetime representing the separation between two events which has an absolute existence independent of the choice of reference frame, and is known as a *four-vector*. This four-vector is known as the displacement four-vector, and represents the displacement in spacetime between the two events E_1 and E_2. In order to distinguish a four-vector from an ordinary three-vector, a superscript arrow will be used.

As was the case with three-vectors, any quantity which transforms in the same way as $\Delta \vec{s}$ is also termed a four-vector. For instance, we have shown that

$$\left.\begin{aligned} E'/c &= \gamma(E/c) - \frac{\gamma v_x}{c} p_x \\ p'_x &= -\frac{\gamma v_x}{c}(E/c) + \gamma p_x \\ p'_y &= p_y \\ p'_z &= p_z \end{aligned}\right\} \tag{5.15}$$

which we can write as

$$\begin{pmatrix} E'/c \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v_x/c & 0 & 0 \\ -\gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix} \tag{5.16}$$

where we see that the same matrix appears on the right hand side as in the transformation law for $\Delta \vec{s}$. This expression relates the components, in two different frames of reference S and S', of the four-momentum of a particle. This four-momentum is, of course, by virtue of this transformation property, also a four-vector. We can note that the ('length')^2 of this four-vector is given by

$$(E/c)^2 - \left[p_x^2 + p_y^2 + p_z^2\right] = (E/c)^2 - \mathbf{p}^2 = (E^2 - \mathbf{p}^2c^2)/c^2 = m_0^2c^2 \tag{5.17}$$

where m_0 is the rest mass of the particle. This quantity is the same (i.e. invariant) in different frames of reference.

A further four-vector is the velocity four-vector

$$\vec{v} \doteq \begin{pmatrix} cdt/d\tau \\ dx/d\tau \\ dy/d\tau \\ dz/d\tau \end{pmatrix} \tag{5.18}$$

where

$$d\tau = ds/c \tag{5.19}$$

and is known as the *proper time interval*. This is the time interval measured by a clock in its own rest frame as it makes its way between the two events an interval ds apart.

To see how the velocity four-vector relates to our usual understanding of velocity, consider a particle in motion relative to the inertial reference frame S. We can identify two events, E_1 wherein the particle is at position (x, y, z) at time t, and a second event E_2 wherein the particle is at $(x + dx, y + dy, z + dz)$ at time $t + dt$. The displacement in space and time between these events will then be represented by the four-vector $d\vec{s}$ defined in Eq. (5.14). Furthermore, during this time interval dt as measured in S, the particle undergoes a displacement $d\mathbf{r} = dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}$ and so has a velocity

$$\mathbf{u} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k} = u_x\mathbf{i} + u_y\mathbf{j} + u_z\mathbf{k}. \tag{5.20}$$

The time interval between the events E_1 and E_2 as measured by a clock moving with the particle will be just the proper time interval $d\tau$ in the rest frame of the particle. We therefore have, by the time dilation formula

$$dt = \frac{d\tau}{\sqrt{1 - (u/c)^2}} \tag{5.21}$$

where u is the speed of the particle. So, if we form the four-velocity to be associated with the two events E_1 and E_2, we write

$$\vec{u} \doteq \begin{pmatrix} cdt/d\tau \\ dx/d\tau \\ dy/d\tau \\ dz/d\tau \end{pmatrix} = \frac{1}{\sqrt{1 - (u/c)^2}} \begin{pmatrix} c \\ dx/dt \\ dy/dt \\ dz/dt \end{pmatrix} = \frac{1}{\sqrt{1 - (u/c)^2}} \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix} \tag{5.22}$$

Thus, if $u << c$, the three spatial components of the four velocity reduces to the usual components of ordinary three-velocity. Note also that the invariant ('length')^2 of the velocity four-vector is just c^2.

Finally, if we take the expression for the four-velocity and multiply by the rest mass of the particle, we get

$$m_0\vec{u} \doteq \frac{1}{\sqrt{1 - (u/c)^2}} \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix} = \begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix} \tag{5.23}$$

which can be recognized as the four momentum defined above.

We can continue in this way, defining four-acceleration

$$\vec{a} = \frac{d\vec{u}}{d\tau} \tag{5.24}$$

and the four-force, also known as the Minkowski force $\vec{F}$:

$$\vec{F} = \frac{d\vec{p}}{d\tau}. \tag{5.25}$$

A direct generalization of the Newtonian definition would have been $\vec{F} = m_0\vec{a}$, but this definition is not applicable to zero rest mass particles, hence the more general alternative in Eq. (5.25).

5.3 Minkowski Space

Till now we have represented a frame of reference S by a collection of clocks and rulers. An alternative way of doing the same thing is to add a fourth axis, the time axis, 'at right angles' to the X, Y, Z axes. On this time axis we can plot the time t that the clock reads at the location of an event. Obviously we cannot draw in such a fourth axis, but we can suppress the Y, Z coordinates for simplicity and draw as in Fig. (5.3):

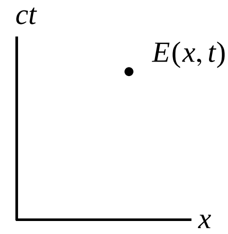

Figure 5.3: An event represented as a point in spacetime.

This representation is known as a spacetime or Minkowski diagram and on it we can plot the positions in space and time of the various events that occur in spacetime. In particular we can plot the motion of a particle through space and time. The curve traced out is known as the world line of the particle. We can note that the slope of such a world line must be greater than the slope of the world line of a photon since all material particles move with speeds less than the speed of light. Some typical world lines are illustrated in Fig. (5.4) below.

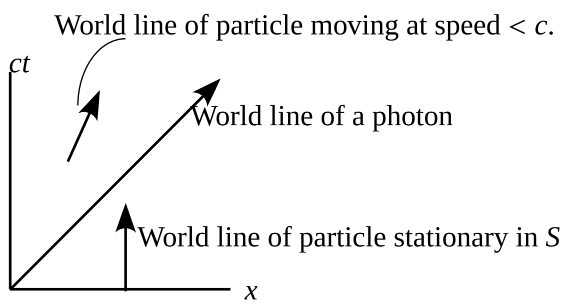

Figure 5.4: Diagram illustrating different kinds of world lines.

The above diagram gives the coordinates of events as measured in a frame of reference S say. We can also use these spacetime diagrams to illustrate Lorentz transformations from one frame of reference to another. Unfortunately, due to the peculiar nature of the interval between two events in spacetime, the new set of axes for some other frame of reference S' is not a simple rotation of the old axes. The equations for the S' axes are determined in a straightforawrd fashion from the Lorentz transformation equations. The x' axis is just the line for which $t' = 0$, which gives

$$ct = \frac{v_x}{c}x \tag{5.26}$$

and for the t' axis, for which $x' = 0$

$$ct = \frac{c}{v_x}x. \tag{5.27}$$

It therefore turns out that these new axes are oblique, as illustrated in Fig. 5.5, and with increasing speeds of S' relative to S, these axes close in on the world line of the photon passing through the common origin.

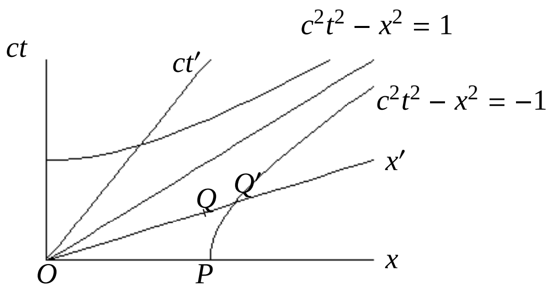

Figure 5.5: Space and time axes for two different reference frames. The rectilinear axes are for the reference frame S, the oblique axes those for a reference frame S' moving with respect to S. The lengths OP and OQ are the same on the figure, but if OP represents a distance of 1 m in S, then OQ' represents the same distance in S'.

It should be noted that in deriving these equations, the γ factor cancels out. But, as this factor plays an integral role in the Lorentz transformation, appearing in both the length contraction and time dilation formulae, it is clear that it is not sufficient to simply determine the new axes in S' if the spacetime diagram is to be used to compare lengths or times in the two reference frames. What also needs to be done is to rescale the units of time and distance along each S' axis. To put it another way, if the two events O and P on the x axis are one metre apart, then two points O and Q on the x' axis which are the same distance apart on the diagram (they are about 1.5 cm in Fig. 5.5) will *not* represent a distance of one metre in S'. To see what separation is required on the x' axis, we proceed as follows.

The spacetime interval between O and P is given by $\Delta s = -1$ m^2. If we now plot all points that have the same spacetime separation from O on this spacetime diagram, we see that these points will lie on the curve

$$(ct)^2 - x^2 = -1 \tag{5.28}$$

which is the equation of a hyperbola. It will cut the x' axis (where $t' = 0$) at the point Q'. But since the interval is the same in all reference frames, we must also have

$$(ct')^2 - x'^2 = -1 \tag{5.29}$$

so at $t' = 0$ we have $x' = 1$. Thus, it is the distance between O and P'' that represents a distance of 1 m in S'. A similar argument can be used to determine the scaling along the time axis in the S' frame, i.e. the point for which $ct' = 1$, where the hyperbola $(ct)^2 - x^2 = 1$ cuts the ct' axis gives the unit of time on the ct' axis.

To illustrate how spacetime diagrams can be used, we will briefly look at length contraction. Consider a rod which is at rest in the reference frame S, one end at O, the other at P, corresponding to a length of 1 m say. The world lines of the ends of the rod will then lie parallel to the ct axis. Now suppose that the position of the ends of the rod is measured at the same time in S', at the time $t' = 0$ in fact. These will be the points O and P' as indicated on Fig. 5.6. It can be seen that OP' is shorter than OQ', the latter being a distance of 1 m in S'. We will not be considering this aspect of spacetime diagrams any further here. However, what we will briefly look at is some of the properties of the spacetime interval Δs that leads to this strange behaviour.

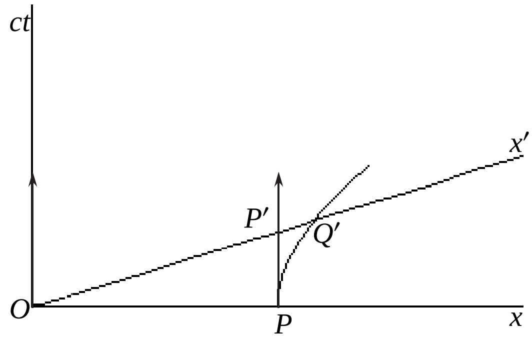

Figure 5.6: A rod of length $OP = 1$ m is stationary in S. Its length in S' is given by OP' which is less than OQ', which has a length of 1 m in S'.

5.4 Properties of Spacetime Intervals

We saw in the preceding section that one of the invariant quantities is the interval Ds as it is just the "length" of the four-vector. As we saw earlier, it is the analogue in spacetime of the familiar distance between two points in ordinary 3-dimensional space. However, unlike the ordinary distance between two points, or more precisely (distance)^2, which is always positive (or zero), the interval between two events E_1 and E_2 i.e. Δs^2, can be positive, zero, or negative. The three different possibilities have their own names:

1. $\Delta s^2 < 0$: E_1 and E_2 are separated by a *space-like* interval.
2. $\Delta s^2 = 0$: E_1 and E_2 are separated by a *light-like* interval.
3. $\Delta s^2 > 0$: E_1 and E_2 are separated by a *time-like* interval.

What these different possibilities represent is best illustrated on a spacetime diagram. Suppose an event O occurs at the spacetime point $(0, 0)$ in some frame of reference S. We can divide the spacetime diagram into two regions as illustrated in the figure below: the shaded region lying between the world lines of photons passing through $(0, 0)$, and the unshaded region lying outside these world lines. Note that if we added a further space axis, in the Y direction say, the world lines of the photons passing through will lie on a cone with its vertex at O. This cone is known as the 'light cone'. Then events such as Q will lie 'inside the light cone', events such as P 'outside the light cone', and events such as R 'on the light cone'.

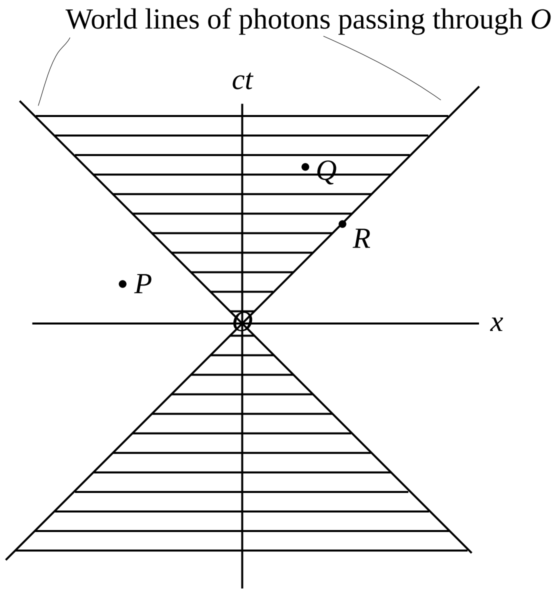

Figure 5.7: The point Q within the light cone (the shaded region) is separated from O by a time-like interval. A signal travelling at a speed less than c can reach Q from O. The point R on the edge of the light cone is separated from O by a light-like interval, and a signal moving at the speed c can reach R from O. The point P is outside the light cone. No signal can reach P from O.

Consider now the sign of Δs^2 between events O and P. Obviously

$$\Delta s^2 = (c\Delta t)^2 - (\Delta x)^2 < 0 \tag{5.30}$$

i.e. all points outside the light cone through O are separated from O by a space-like interval. Meanwhile, for the event Q we have

$$\Delta s^2 = (c\Delta t)^2 - (\Delta x)^2 > 0 \tag{5.31}$$

i.e. all points inside the light cone through O are separated from O by a time-like interval. Finally for R we have

$$\Delta s^2 = (c\Delta t)^2 - (\Delta x)^2 = 0 \tag{5.32}$$

i.e. all points on the light cone through O are separated from O by a light-like interval.

The physical meaning of these three possibilities can be seen if we consider whether or not the event O can in some way affect the events P, Q, or R. In order for one event to physically affect another some sort of signal must make its way from one event to the other. This signal can be of any kind: a flash of light created at O, a massive particle emitted at O, a piece of paper with a message on it and placed in a bottle. Whatever it is, in order to be present at the other event and hence to either affect it (or even to cause it) this signal must travel the distance Δx in time Δt, i.e. with speed $\Delta x/\Delta t$.

We can now look at what this will mean for each of the events P, Q, R. Firstly, for event P we find from Eq. (5.30) that $\Delta x/\Delta t > c$. Thus the signal must travel faster than the speed of light, which is not possible. Consequently event O cannot affect, or cause event P. Secondly, for event Q we find from Eq. (5.31) that $\Delta x/\Delta t < c$ so the signal will travel at a speed less than the speed of light, so event O can affect (or cause) event Q. Finally, for R we find from Eq. (5.32) that $\Delta x/\Delta t = c$ so that O can effect R by means of a signal travelling at the speed of light. In summary we can write

1. Two events separated by a space-like interval cannot affect one another;

2. Two events separated by a time-like or light-like interval can affect one another.

Thus, returning to our spacetime diagram, we have:

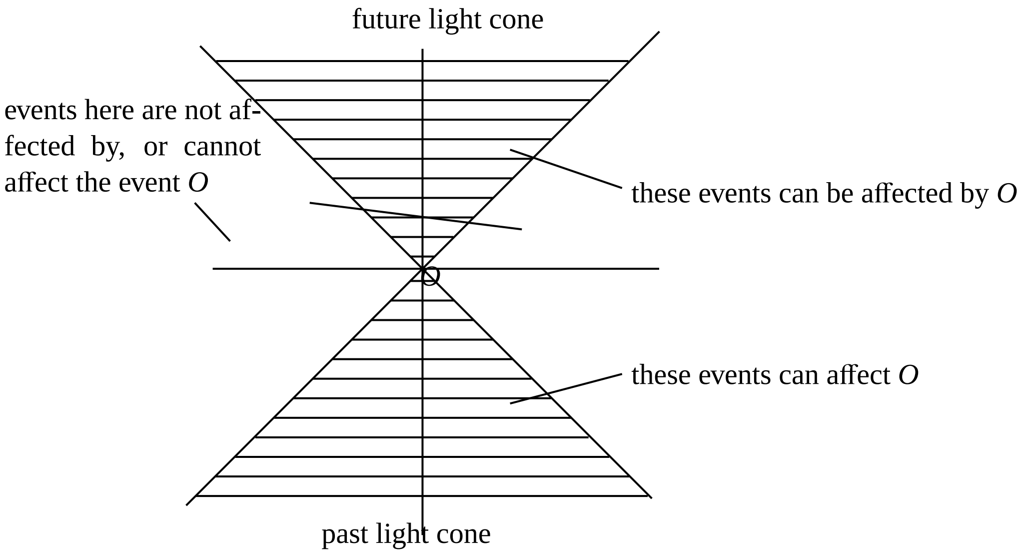

Figure 5.8: Future and past light cones of the event O

All the events that can be influenced by O constitute the future of event O while all events that can influence O constitute the past of event O.

5.5 Four-Vector Notation

It is at this point that a diversion into further mathematical development of the subject is necessary. For the present, we will be more concerned with the *way* that the physics is described mathematically, rather than the content of the physics itself. This is necessary to put in place the notation and mathematical machinery that is used in general relativity (and in further developments in special relativity, for that matter.) The first step in this direction is to introduce a more uniform way of naming the components of the four-vector quantities introduced above which better empahsizes its vector nature, that is:

$$x^0 = ct, \qquad x^1 = x, \qquad x^2 = y, \qquad x^3 = z \tag{5.33}$$

where the superscript numbers are NOT powers of x. In the same way, the components of the momentum four-vector will be

$$p^0 = E/c, \qquad p^1 = p_x, \qquad p^2 = p_y, \qquad p^3 = p_z \tag{5.34}$$

and similarly for other four-vectors.

In terms of these names for the components we can write the Lorentz transformation equations as

$$(\Delta x^\mu)' = \sum_{\nu=0}^{3} \Lambda^\mu_\nu \Delta x^\nu \tag{5.35}$$

where, if S' is moving with velocity v_x relative to S, then the Λ^μ_ν will be the components of the 4×4 matrix appearing in Eq. (5.10) and Eq. (5.16), that is[1]

$$\Lambda^\mu_\nu = \begin{pmatrix} \Lambda^0_0 & \Lambda^0_1 & \Lambda^0_2 & \Lambda^0_3 \\ \Lambda^1_0 & \Lambda^1_1 & \Lambda^1_2 & \Lambda^1_3 \\ \Lambda^2_0 & \Lambda^2_1 & \Lambda^2_2 & \Lambda^2_3 \\ \Lambda^3_0 & \Lambda^3_1 & \Lambda^3_2 & \Lambda^3_3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v_x/c & 0 & 0 \\ -\gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.36}$$

It is at this point that we make the first of two notational changes. First we note that we have usually represented quantities as observed in S' by attaching a prime to the symbol, e.g. x', t' and so on. Now, we will attach the prime to the index, so that we will henceforth write:

$$\Delta x^{\mu'} = \sum_{\nu=0}^{3} \Lambda^{\mu'}_\nu \Delta x^\nu \tag{5.37}$$

where now $\mu' = 0', 1', 2'$ or $3'$, so that the transformation matrix is now

$$\Lambda^{\mu'}_\nu = \begin{pmatrix} \Lambda^{0'}_0 & \Lambda^{0'}_1 & \Lambda^{0'}_2 & \Lambda^{0'}_3 \\ \Lambda^{1'}_0 & \Lambda^{1'}_1 & \Lambda^{1'}_2 & \Lambda^{1'}_3 \\ \Lambda^{2'}_0 & \Lambda^{2'}_1 & \Lambda^{2'}_2 & \Lambda^{2'}_3 \\ \Lambda^{3'}_0 & \Lambda^{3'}_1 & \Lambda^{3'}_2 & \Lambda^{3'}_3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v_x/c & 0 & 0 \\ -\gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.38}$$

It is important to recognize that this matrix, as used in Eq. (5.37), transforms the components of 4-vectors in S to the components in S'. If we were to carry out a transformation from S' to S, we would have to write

$$\Delta x^{\mu} = \sum_{\nu'=0'}^{3'} \Lambda^{\mu}_{\nu'} \Delta x^{\nu'} \tag{5.39}$$

with now

$$\Lambda^{\mu}_{\nu'} = \begin{pmatrix} \Lambda^0_{0'} & \Lambda^0_{1'} & \Lambda^0_{2'} & \Lambda^0_{3'} \\ \Lambda^1_{0'} & \Lambda^1_{1'} & \Lambda^1_{2'} & \Lambda^1_{3'} \\ \Lambda^2_{0'} & \Lambda^2_{1'} & \Lambda^2_{2'} & \Lambda^2_{3'} \\ \Lambda^3_{0'} & \Lambda^3_{1'} & \Lambda^3_{2'} & \Lambda^3_{3'} \end{pmatrix} = \begin{pmatrix} \gamma & \gamma v_x/c & 0 & 0 \\ \gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.40}$$

where we note that $v_x \to -v_x$ as we are transforming 'the other way', that is, from S' to S. It is reasonable to expect that the two matrices for the Lorentz transformations from S to S' and vice versa would be inverses of each other. That this is indeed the case can be readily confirmed by multiplying the two matrices together. This point is discussed further in the following Section.

The second change in notation is very important as it offers considerable simplification of what would otherwise be exceedingly complicated expressions. This new notation goes under the name of the Einstein summation convention.

5.5.1 The Einstein Summation Convention

This convention, which Einstein looked on as his greatest invention, means replacing sums like

$$\Delta x^{\mu'} = \sum_{\nu=0}^{3} \Lambda^{\mu'}_\nu \Delta x^\nu$$

[1]Note the unusual notation in which the symbol for an *element* of a matrix is used as the symbol for the *complete* matrix. This is done as a shorthand convenience, and though mathematically inelegant, it does have its uses.

by

$$\Delta x^{\mu'} = \Lambda^{\mu'}_{\nu} \Delta x^{\nu} \tag{5.41}$$

with the understanding that whenever there is any repeated (greek) index appearing in a 'one up', 'one down' combination the summation over the four values of the repeated index is understood. Thus here, as the index ν appears 'down' in $\Lambda^{\mu'}_{\nu}$ and 'up' in Δx^{ν}, a summation over this index is understood.

There are a number of important features associated with the convention.

Dummy indices: A repeated index is known as a dummy index, by which is meant that any (greek) symbol can be used instead without a change in meaning, i.e.

$$\Delta x^{\mu'} = \Lambda^{\mu'}_{\alpha} \Delta x^{\alpha} = \Lambda^{\mu'}_{\beta} \Delta x^{\beta} = \ldots. \tag{5.42}$$

This changing around of dummy indices can be a useful trick in simplifying expressions, particularly when a substitution has to be made, as will be illustrated later.

No summation implied: If an index is repeated, but both occur in an up position or in a down position, then no summation is implied, i.e.

$$\Gamma^{\mu\mu} \neq \Gamma^{00} + \Gamma^{11} + \Gamma^{22} + \Gamma^{33}.$$

No meaning assigned: If an index is repeated more than twice, then no meaning is assigned to such a combination, i.e. $\Gamma^{\mu\mu}_{\mu}$ does not have an unambiguous meaning. If such a combination should occur, then there is a good chance that an error has been made!

Free index: Any index that is not repeated in a one up one down arrangement is known as a free index – we are free to give it any of its four possible values. In an equation, all free indices must appear on both sides of the equation in the same i.e. up or down, position. Thus

$$a_{\mu} = \Lambda^{\mu}_{\nu'} a^{\nu'}$$

is NOT correct, as μ appears in different positions on either side of the equation. The following example is also not correct

$$G = g_{\alpha\beta} u^{\beta}.$$

as the free index α appears only on the right hand side of the equation.

The name of a free index can be changed, of course, provided it is changed on both sides of an equation. Changing the name of a free index is also a useful trick when manipulating expressions, particularly when one expression is to be substituted into another.

Multiple repeated indices: If more than one pair of repeated indices occurs, then a summation is implied over *all* the repeated indices i.e.

$$\begin{aligned} g_{\mu\nu} a^{\mu} b^{\nu} &= g_{0\nu} a^0 b^{\nu} + g_{1\nu} a^1 b^{\nu} + g_{2\nu} a^2 b^{\nu} + g_{3\nu} a^3 b^{\nu} \\ &= g_{00} a^0 b^0 + g_{01} a^0 b^1 + g_{02} a^0 b^2 + g_{03} a^0 b^3 \\ &\quad + g_{10} a^1 b^0 + g_{11} a^1 b^1 + g_{12} a^1 b^2 + g_{13} a^1 b^3 \\ &\quad + g_{20} a^2 b^0 + g_{21} a^2 b^1 + g_{22} a^2 b^2 + g_{23} a^2 b^3 \\ &\quad + g_{30} a^3 b^0 + g_{31} a^3 b^1 + g_{32} a^3 b^2 + g_{33} a^3 b^3. \end{aligned}$$

To see the convention in action, we will use it to show that the transformations associated with the matrices $\Lambda^{\mu'}_{\nu}$ and $\Lambda^{\nu}_{\beta'}$ are inverses of each other. We begin with the equation

$$\Delta x^{\mu'} = \Lambda^{\mu'}_{\nu}\Delta x^{\nu}. \tag{5.43}$$

which represents a transformation from S to S', and the inverse of this equation

$$\Delta x^{\nu} = \Lambda^{\nu}_{\beta'}\Delta x^{\beta'} \tag{5.44}$$

which represents a transformation from S' to S. We can substitute this into Eq. (5.43) to give

$$\Delta x^{\mu'} = \Lambda^{\mu'}_{\nu}\Lambda^{\nu}_{\beta'}\Delta x^{\beta'}. \tag{5.45}$$

Now define a new quantity

$$\delta^{\mu'}_{\beta'} = \Lambda^{\mu'}_{\nu}\Lambda^{\nu}_{\beta'} \tag{5.46}$$

which we can identify as the (μ', β') element of the product of the matrices $\Lambda^{\mu'}_{\nu}$ and $\Lambda^{\nu}_{\beta'}$. In fact, we see that $\Lambda^{\mu'}_{\nu}$ is the inverse of $\Lambda^{\mu}_{\beta'}$, i.e.

$$\begin{pmatrix} \gamma & -\gamma v_x/c & 0 & 0 \\ -\gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \gamma & \gamma v_x/c & 0 & 0 \\ \gamma v_x/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.47}$$

so that $\delta^{\mu'}_{\beta'}$ are just the components of the identity matrix, i.e.

$$\begin{aligned} \delta^{\mu'}_{\beta'} &= 1 \quad \mu' = \beta' \\ &= 0 \quad \mu' \neq \beta' \end{aligned} \tag{5.48}$$

The quantity defined by Eq. (5.48) is known as the Kronecker delta. It has the unique property of having the same values in any reference frame, i.e. in S we have

$$\begin{aligned} \delta^{\mu}_{\beta} &= 1 \quad \mu = \beta \\ &= 0 \quad \mu \neq \beta \end{aligned}$$

as can be confirmed by evaluating $\delta^{\mu}_{\beta} = \lambda^{\mu}_{\mu'}\Lambda^{\beta'}_{\beta}\delta^{\mu'}_{\beta'}$.

5.5.2 Basis Vectors and Contravariant Components

The four components of the spacetime displacement four-vector $\Delta\vec{s}$ can be used to construct the four-vector itself by introducing a set of basis vectors. Thus we will write our displacement vector $\Delta\vec{s}$ as

$$\Delta\vec{s} = \Delta x^{\mu}\vec{e}_{\mu}. \tag{5.49}$$

It is tempting to think of $\vec{e}_i$, $i = 1, 2, 3$ as being, in effect, the usual unit vectors in 3-space, that is $\hat{i}$, $\hat{j}$, and $\hat{k}$ respectively, but this is not tenable when we consider how these basis vectors transform between different reference frames. What is found is that even if a reference frame S' is moving in the x direction relative to S, in which case it might be expected that $\vec{e}_{1'}$ would still 'point' in the $\hat{i}$ direction, it is found that $\vec{e}_{1'} \neq \vec{e}_1$ as it acquires components in the 'time' direction. Thus we must consider these basis vectors as being abstract vectors in spacetime that may happen to coincide with the familiar unit vectors under some circumstances.

Any four-vector such as the velocity, momentum or Minkowski force four-vectors can be expressed in terms of these basis vectors in the same way, e.g.

$$\vec{p} = p^{\mu}\vec{e}_{\mu} \tag{5.50}$$

but we will develop the ideas here in terms of the spacetime displacement four-vector $\Delta\vec{s}$.

We have said repeatedly that $\Delta\vec{s}$ is a geometrical object that exists in spacetime independent of any choice of reference frame with which to assign its components, so we can equally well write the above expression for $\Delta\vec{s}$ as

$$\Delta\vec{s} = \Delta x^{\mu'}\vec{e}_{\mu'} \tag{5.51}$$

where $\vec{e}_{\mu'}$ are the new basis vectors in S'. If we now use the Lorentz transformation to write

$$\Delta x^{\mu'} = \Lambda^{\mu'}_{\nu}\Delta x^{\nu} \tag{5.52}$$

we then get

$$\Delta\vec{s} = \Lambda^{\mu'}_{\nu}\Delta x^{\nu}\vec{e}_{\mu'} = \Delta x^{\nu}\Lambda^{\mu'}_{\nu}\vec{e}_{\mu'} = \Delta x^{\nu}\vec{e}_{\nu} \tag{5.53}$$

where now

$$\vec{e}_{\nu} = \Lambda^{\mu'}_{\nu}\vec{e}_{\mu'}. \tag{5.54}$$

Equivalently, we have

$$\vec{e}_{\nu'} = \Lambda^{\mu}_{\nu'}\vec{e}_{\mu} \tag{5.55}$$

so that, for instance

$$\vec{e}_{1'} = \Lambda^{\mu}_{1'}\vec{e}_{\mu} = \gamma\frac{v_x}{c}\vec{e}_0 + \gamma\vec{e}_1 \tag{5.56}$$

which shows, as was intimated above, that $\vec{e}_{1'}$ is a linear combination of $\vec{e}_1$ and $\vec{e}_0$ – i.e. it does not 'point' in the same direction as $\vec{e}_1$, even though S and S' are moving in the x direction relative to one another.

If we now compare the two results

$$\vec{e}_{\mu} = \Lambda^{\nu'}_{\mu}\vec{e}_{\nu'}$$

and

$$\Delta x^{\mu} = \Lambda^{\mu}_{\nu'}\Delta x^{\nu'} \tag{5.57}$$

we see that the first involves the transformation matrix elements $\Lambda^{\nu'}_{\mu}$, the second involves the elements $\Lambda^{\mu}_{\nu'}$ of the inverse matrix (see Eq. (5.47)). Thus the basis vectors and the components transform in 'opposite ways' – they do so in order to guarantee that the interval is the same in all reference frames. Because of this contrary way of transforming, the components Δx^{μ} are referred to as the *contravariant* components of $\Delta\vec{s}$.

In a corresponding way, the components of other four-vectors, such as the components p^{μ} of the momentum four-vector $\vec{p} = p^{\mu}\vec{e}_{\mu}$ will be understood as being contravariant components.

5.5.3 The Metric Tensor

Having defined the spacetime displacement four-vector $\Delta\vec{s}$, we can proceed to define its length in the usual way. But first, we make yet another minor change in notation, namely that we now write Δs^2 rather than $(\Delta s)^2$. Thus, we have

$$\begin{aligned}\Delta s^2 &= \Delta\vec{s}\cdot\Delta\vec{s}\\ &= \left(\Delta x^0\vec{e}_0 + \Delta x^1\vec{e}_1 + \Delta x^2\vec{e}_2 + \Delta x^3\vec{e}_3\right)\cdot\left(\Delta x^0\vec{e}_0 + \Delta x^1\vec{e}_1 + \Delta x^2\vec{e}_2 + \Delta x^3\vec{e}_3\right)\\ &= \left(\Delta x^0\right)^2 - \left(\Delta x^1\right)^2 - \left(\Delta x^2\right)^2 - \left(\Delta x^3\right)^2.\end{aligned} \tag{5.58}$$

Hence we must conclude that

$$\left.\begin{aligned} \vec{e}_\mu \cdot \vec{e}_\nu &= 0 \qquad \mu \neq \nu \\ \vec{e}_0 \cdot \vec{e}_0 &= 1 \\ \vec{e}_i \cdot \vec{e}_i &= -1 \quad i = 1, 2, 3 \end{aligned}\right\} \tag{5.59}$$

so they are most unusual basis vectors indeed!

At this point we introduce a new quantity

$$g_{\mu\nu} = \vec{e}_\mu \cdot \vec{e}_\nu \tag{5.60}$$

which, written as a matrix, looks like

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{5.61}$$

In terms of this quantity, the interval Δs can be written

$$\Delta s^2 = g_{\mu\nu} \Delta x^\mu \Delta^\nu. \tag{5.62}$$

The quantity $g_{\mu\nu}$ plays a central role in defining the geometrical properties of spacetime in that in curved spacetime the components of $g_{\mu\nu}$ are not simple constants but rather are functions of the spacetime coordinates x^μ. A more precise statement is that in the presence of curvature, no matter what frame of reference we use to describe the events in spacetime, there are none for which all the $g_{\mu\nu}$ are constants given by Eq. (5.61) throughout all spacetime. In the particular case in which the $g_{\mu\nu}$ have the constant values given in Eq. (5.61), then spacetime is said to be *flat*. In that case, a different notation is occasionally used, that is $g_{\mu\nu}$ is written $\eta_{\mu\nu}$. As it plays a role in determining the interval or 'distance' between two events in spacetime, $g_{\mu\nu}$ is referred to as the *metric* tensor. Why it is called a tensor is something to be examined later.

There are two properties of $g_{\mu\nu}$ that are worth keeping in mind. First, it is symmetric in the indices, i.e.

$$g_{\mu\nu} = g_{\nu\mu} \tag{5.63}$$

and secondly, it has the same components in all reference frames, i.e.

$$g_{\mu'\nu'} = \Lambda^\alpha_{\mu'} \Lambda^\beta_{\nu'} g_{\alpha\beta} \tag{5.64}$$

a result that can be confirmed by direct calculation.

5.5.4 Covectors and Covariant Components

The metric tensor plays another useful role in that we can define a new set of quantities

$$\Delta x_\mu = g_{\mu\nu} \Delta x^\nu, \tag{5.65}$$

a procedure known as 'lowering an index'. Using the values of $g_{\mu\nu}$ (or $\eta_{\mu\nu}$) given in Eq. (5.59) we can easily evaluate these quantities:

$$\left.\begin{aligned} \Delta x_0 &= \Delta x^0 \\ \Delta x_i &= -\Delta x^i, \quad i = 1, 2, 3. \end{aligned}\right\} \tag{5.66}$$

In terms of the Δx_μ, the interval becomes

$$\Delta s^2 = g_{\mu\nu}\Delta x^\mu \Delta x^\nu = \Delta x_\mu \Delta x^\mu. \tag{5.67}$$

If we write this out in matrix form we get

$$\Delta s^2 = \begin{pmatrix}\Delta x_0 & \Delta x_1 & \Delta x_2 & \Delta x_3\end{pmatrix}\begin{pmatrix}\Delta x^0\\ \Delta x^1\\ \Delta x^2\\ \Delta x^3\end{pmatrix}. \tag{5.68}$$

The column vector here represents the four-vector $\Delta\vec{s}$, but what does the row vector represent? The fact that it is written out as a row vector, and its components are different from those that appear in the column vector, suggests that it represents a different mathematical object as compared to the four-vector $\Delta\vec{s}$, and so we (temporarily) give it a new name, $\Delta\tilde{s}$:

$$\Delta\tilde{s} \doteq \begin{pmatrix}\Delta x_0 & \Delta x_1 & \Delta x_2 & \Delta x_3\end{pmatrix}. \tag{5.69}$$

The newly defined mathematical object is known as a one-form or a covector with components Δx_μ. We can define a set of basis covectors, $\tilde{e}_\mu$ so that we have

$$\Delta\tilde{s} = \Delta x_\mu \tilde{e}^\mu \tag{5.70}$$

but as we will soon see, we will not need to develop this idea any further.

The components Δx_μ will, of course, be different in different reference frames. We can derive the transformation law by, once again, making use of the fact that Δs^2 is the same in all reference frames to write

$$\Delta s^2 = \Delta x_{\mu'}\Delta x^{\mu'}. \tag{5.71}$$

Using $\Delta x^{\mu'} = \Lambda^{\mu'}_\nu \Delta x^\nu$, this becomes

$$\Delta s^2 = \Lambda^{\mu'}_\nu \Delta x_{\mu'}\Delta x_\nu = \Delta x_\nu \Delta x^\nu \tag{5.72}$$

where

$$\Delta x_\nu = \Lambda^{\mu'}_\nu \Delta x_{\mu'}. \tag{5.73}$$

If we compare this with Eq. (5.54), that is $\vec{e}_\nu = \Lambda^{\mu'}_\nu \vec{e}_{\mu'}$ we see that Δx_ν and $\vec{e}_\nu$ transform in exactly the same way. Consequently, the Δx_ν are referred to as the *covariant* components of $\Delta\tilde{s}$.

We have arrived at a state of affairs analogous to what we have in quantum mechanics, namely that

$$\Delta\vec{s} \to |\psi\rangle \quad \text{and} \quad \Delta\tilde{s} \to \langle\psi|$$

though here, no complex conjugation is required as the components of $\Delta\vec{s}$ are all real. Further, just as in quantum mechanics we can equally well describe the state of a physical system in terms of a bra or ket vector, we have here a perfect one-to-one correspondence between $\Delta\vec{s}$ and $\Delta\tilde{s}$. In fact, in general, no distinction need be drawn between them as they equally well represent the same geometrical object in spacetime, so in future we will have no need to talk about the covector $\Delta\tilde{s}$, and instead will simply refer to the four-vector $\Delta\vec{s}$ which has covariant components Δx_μ or contravariant components Δx^μ.

Any four-vector, such as the velocity, acceleration, and Minkowski force four-vectors can be expressed in terms of its covariant components in the same way as the spacetime displacement vector, with the components transforming in exactly the same fashion as in Eq. (5.73). In all cases, the covariant components of these four-vectors will be related to their contravariant counterparts in the same way as for the components of $\Delta\vec{s}$, e.g. for the momentum four-vector:

$$\left.\begin{aligned} p_0 &= p^0 \\ p_i &= -\,p^i, \quad i = 1,2,3. \end{aligned}\right\} \tag{5.74}$$

5.5.5 Transformation of Differential Operators

To see an important example of a situation in which the covariant components of a four-vector naturally arise, consider the derivatives

$$\frac{\partial\phi}{\partial x^\mu}$$

where ϕ is some function that has the same value in all reference frames (a scalar function). To transform these derivatives to their values in another reference frame, we need to use the chain rule for partial derivatives, i.e.

$$\frac{\partial\phi}{\partial x^{0'}} = \frac{\partial\phi}{\partial x^0}\frac{\partial x^0}{\partial x^{0'}} + \frac{\partial\phi}{\partial x^1}\frac{\partial x^1}{\partial x^{0'}} + \frac{\partial\phi}{\partial x^2}\frac{\partial x^2}{\partial x^{0'}} + \frac{\partial\phi}{\partial x^3}\frac{\partial x^3}{\partial x^{0'}}. \tag{5.75}$$

Using the Lorentz transformation

$$x^\mu = \Lambda^\mu_{\nu'} x^{\nu'} \tag{5.76}$$

we see that

$$\frac{\partial x^0}{\partial x^{0'}} = \Lambda^0_{0'}. \tag{5.77}$$

If we carry out the same calculation for all the partial derivatives, we find that

$$\frac{\partial x^\mu}{\partial x^{\nu'}} = \Lambda^\mu_{\nu'} \tag{5.78}$$

so that

$$\frac{\partial\phi}{\partial x^{\nu'}} = \Lambda^\mu_{\nu'}\frac{\partial\phi}{\partial x^\mu}. \tag{5.79}$$

If we introduce a new notation and write

$$\frac{\partial\phi}{\partial x^\nu} = \partial_\nu\phi \tag{5.80}$$

then Eq. (5.79) becomes

$$\partial_{\nu'}\phi = \Lambda^\mu_{\nu'}\partial_\mu\phi \tag{5.81}$$

which is just the transformation rule for covariant components, Eq. (5.73) of a four-vector. In fact, it is usual practice to treat the differential operators ∂_μ themselves as being the covariant components of a four-vector, and write

$$\partial_{\nu'} = \Lambda^\mu_{\nu'}\partial_\mu. \tag{5.82}$$

5.6 Tensors

The last formal mathematical tool that we need to introduce is the concept of a tensor. A tensor is a generalization of the idea of a four-vector, and as such a tensor represents geometrical object existing in spacetime, but one that is even more difficult to visualize than a four-vector.

One viewpoint with regard to tensors is that they can be considered as being 'operators' that act upon four-vectors to produce real numbers, and that is the way that the concept will be introduced here. The connection between tensors defined in this manner, and concepts already introduced will emerge later, as will the physical applications of the idea.

Thus, we begin with a definition.

> *A tensor* $\mathsf{T}(\vec{a},\vec{b},\vec{c},\ldots)$ *is a linear function of the four-vectors* $\vec{a},\vec{b},\vec{c},\ldots$ *that maps these four-vectors into the real numbers.*

Different rules for how the real number is calculated from the vector arguments then gives rise to different tensors. The manner of definition, namely that no mention is made of any reference frame, means that a tensor is a quantity that is independent of the choice of reference frame.

The following properties and definitions are important:

Rank The *rank* of a tensor is the number of vector arguments. Thus a tensor of rank 1 will be the function of one vector only, i.e. $\mathsf{p}(\vec{a})$ will define a tensor of rank one, $\mathsf{g}(\vec{a}, \vec{b})$, a tensor of rank 2 and so on.

Linearity That the function $\mathsf{T}(\vec{a}, \vec{b}, \vec{c}, \ldots)$ is linear means that for any numbers u and v

$$\mathsf{T}(u\vec{a} + v\vec{b}, \vec{c}, \ldots) = u\mathsf{T}(\vec{a}, \vec{c}, \ldots) + v\mathsf{T}(\vec{b}, \vec{c}, \ldots) \tag{5.83}$$

with same being true for all the arguments, i.e.

$$\mathsf{T}(\vec{a}, u\vec{b} + v\vec{c}, \ldots) = u\mathsf{T}(\vec{a}, \vec{b}, \ldots) + v\mathsf{T}(\vec{a}, \vec{c}, \ldots). \tag{5.84}$$

Tensor Components The components of a tensor are the values of the tensor obtained when evaluated for the vectors $\vec{a}$, $\vec{b}$, $\vec{c}$, $\ldots$ equal to the basis vectors. Thus, we have

$$T_{\mu\nu\alpha\ldots} = \mathsf{T}(\vec{e}_\mu, \vec{e}_\nu, \vec{e}_\alpha, \ldots). \tag{5.85}$$

As a consequence of this and the linearity of T, we have

$$\mathsf{T}(\vec{a}, \vec{b}, \vec{c}, \ldots) = \mathsf{T}(\vec{e}_\mu, \vec{e}_\nu, \vec{e}_\alpha, \ldots)a^\mu b^\nu c^\alpha \ldots = T_{\mu\nu\alpha\ldots}a^\mu b^\nu c^\alpha \ldots. \tag{5.86}$$

Raising and Lowering Indices The process of raising and lowering indices can be carried through with the components of a tensor in the expected way. Thus, we can write

$$T^\mu{}_{\nu\alpha\ldots} = g^{\beta\mu}T_{\beta\nu\alpha\ldots} \tag{5.87}$$

or

$$T_\mu{}^\nu{}_\alpha \ldots = g^{\beta\nu}T_{\mu\beta\alpha\ldots}. \tag{5.88}$$

Corresponding to this we would have, for instance

$$\mathsf{T}(\vec{a}, \vec{b}, \vec{c}, \ldots) = T_{\mu\nu\alpha\ldots}a^\mu b^\nu c^\alpha \ldots = T_{\mu\nu\alpha\ldots}g^{\beta\mu}a_\beta b^\nu c^\alpha \ldots. \tag{5.89}$$

Where the implied sum over β means that we are applying the raising procedure to a_β. But, if we reqroup the terms, we have

$$\mathsf{T}(\vec{a}, \vec{b}, \vec{c}, \ldots) = g^{\beta\mu}T_{\mu\nu\alpha\ldots}a_\beta b^\nu c^\alpha \ldots \tag{5.90}$$

where we now see that the implied sum over μ means that we are raising an index in the components of the tensor, i.e.

$$\mathsf{T}(\vec{a}, \vec{b}, \vec{c}, \ldots) = T^\beta{}_{\nu\alpha\ldots}a_\beta b^\nu c^\alpha \ldots \tag{5.91}$$

This flexibility in moving indices up and down by the application of $g^{\mu\nu}$ or $g_{\mu\nu}$ means that we can express the components of any tensor in a number of ways that differ by the position of the indices. The different ways in which this is done is described by different terminology, i.e.

$T_{\mu\nu\alpha\ldots}$	covariant components of T
$T_\mu{}^\nu{}_{\alpha\ldots}$ or $T^{\mu\nu}{}_{\alpha\ldots}$ and other combinations of up and down indices	mixed components of T
$T^{\mu\nu\alpha\ldots}$	contravariant components of T.

Being able to raise and lower indices of a tensor raises the possibility of introducing a further mathematical manipulation of tensors. We will illustrate it in the case of a tensor of rank 2, $\mathsf{T}(\vec{a}, \vec{b})$ with covariant components $T_{\mu\nu}$ and mixed components $T_\mu{}^\beta$ where

$$T_\mu{}^\beta = g^{\beta\nu} T_{\mu\nu}. \tag{5.92}$$

If we set $\mu = \beta$ in the tensor component $T_\mu{}^\beta$, we obtain $T_\mu{}^\mu$ which, according to the summation convention implies a sum must be taken over the repeated index μ i.e.

$$T_\mu{}^\mu = T_0{}^0 + T_1{}^1 + T_2{}^2 + T_3{}^3. \tag{5.93}$$

This procedure is known as a *contraction* of the tensor. The effect of contraction of a tensor is to lower the rank of the tensor by 2, as seen here where the result is a number (a scalar), a tensor of rank 0.

5.6.1 Some Examples

At this point it is useful to introduce an example, a second rank tensor $\mathsf{g}(\vec{a}, \vec{b})$ defined by

$$\mathsf{g}(\vec{a}, \vec{b}) = \vec{a} \cdot \vec{b}. \tag{5.94}$$

It is clearly the case that

$$\mathsf{g}(\vec{e}_\mu, \vec{e}_\nu) = \vec{e}_\mu \cdot \vec{e}_\nu = g_{\mu\nu} \tag{5.95}$$

which is just the elements of the metric tensor, introduced in Section 5.5.3. For an arbitrary pair of vectors $\vec{a}$ and $\vec{b}$, this becomes

$$\mathsf{g}(\vec{a}, \vec{b}) = g_{\mu\nu} a^\mu b^\nu \tag{5.96}$$

so that, in particular

$$\mathsf{g}(\Delta\vec{s}, \Delta\vec{s}) = g_{\mu\nu} \Delta x^\mu \Delta x^\nu = \Delta s^2 \tag{5.97}$$

which is just the expression for the interval.

A closely related example is that of a rank 1 tensor, i.e. $\mathsf{a}(\vec{b})$:

$$\mathsf{a}(\vec{b}) = \mathsf{a}(b^\nu \vec{e}_\nu) = \mathsf{a}(\vec{e}_\nu) b^\nu = a_\nu b^\nu = g_{\mu\nu} a^\mu b^\nu = \vec{a} \cdot \vec{b} \tag{5.98}$$

where we have identified the components $\mathsf{a}(\vec{e}_\nu) = a_\mu$ of the tensor as the covariant components of a vector $\vec{a}$. In other words, a tensor or rank 1 is identical to a four-vector.

Tensors of rank zero have no components: they are known as scalars in the same way as quantities without vector components in ordinary three space are referred to as scalars.

5.6.2 Transformation Properties of Tensors

The last property of tensors that we need to consider is the manner in which they transform between different reference frames. This can be derived in a direct fashion that makes use of the fact that the tensors were defined in a way that is independent of the choice of reference frame, and hence tensors are geometrical objects (in the same way as four-vectors are) that have an existence in spacetime independent of any choice of reference frame. With that in mind, we can immediately write for the covariant components of a tensor T

$$\mathsf{T}(\vec{a}, \vec{b}, \vec{c}, \ldots) = T_{\mu\nu\alpha\ldots} a^\mu b^\nu c^\alpha \ldots = T_{\mu'\nu'\alpha'\ldots} a^{\mu'} b^{\nu'} c^{\alpha'} \ldots. \tag{5.99}$$

By using $a^{\mu'} = \Lambda^{\mu'}_{\ \nu} a^\nu$ and similarly for the other vector components, this becomes

$$T_{\mu\nu\alpha\ldots} a^\mu b^\nu c^\alpha \ldots = T_{\mu'\nu'\alpha'\ldots} \Lambda^{\mu'}_{\ \mu} \Lambda^{\nu'}_{\ \nu} \Lambda^{\alpha'}_{\ \alpha} \ldots a^\mu b^\nu c^\alpha \ldots \tag{5.100}$$

As the vectors $\vec{a}, \vec{b}, \vec{c} \ldots$ are arbitrary, we have

$$T_{\mu\nu\alpha\ldots} = \Lambda^{\mu'}_{\ \mu} \Lambda^{\nu'}_{\ \nu} \Lambda^{\alpha'}_{\ \alpha} \ldots T_{\mu'\nu'\alpha'\ldots} \tag{5.101}$$

In other words, the transformation is carried out in the same fashion as we have seen for the single index case (i.e. for the components of vectors). In a similar way (by use of $g^{\mu\nu}$ to raise indices), we can show for the contravariant components that

$$T^{\mu\nu\alpha\ldots} = \Lambda^{\mu}_{\ \mu'} \Lambda^{\nu}_{\ \nu'} \Lambda^{\alpha}_{\ \alpha'} \ldots T^{\mu'\nu'\alpha'\ldots} \tag{5.102}$$

The results Eq. (5.101) and Eq. (5.102), and a corresponding result for mixed components of the tensor T can be used as a test to see whether or not a multi-indexed quantity is, in fact, a tensor. We shall see how this can be implemented in the case of the Faraday tensor used to describe the electromagnetic field.

Chapter 6

Electrodynamics in Special Relativity

ONE of the driving forces behind Einstein's formulation of the principles of special relativity was the deep significance he attached to the laws of electromagnetism. It is therefore not too surprising to find that these laws can be expressed in the language of four-vectors and tensors in a way that explicitly shows that electromagnetism is consistent with the principles of special relativity. The central feature of this relativistic formulation of Maxwell's theory is the Faraday tensor.

6.1 The Faraday Tensor

By judicious arguments based on applying length contraction and time dilation arguments to the various basic laws of electromagnetism: Ampere's law for the magnetic field produced by currents, Faraday's law of magnetic induction for the electric fields produced by a time varying magentic field, and Gauss's law for the electric field produced by static electric charges – all of which are expressed in terms of either line or surface integrals, or else by working directly from Maxwell's equations (which are simply restatements of the integral laws in differential form), it is possible to show that electric and magnetic fields $\mathbf{E}(x,y,z,t)$ and $\mathbf{B}(x,y,z,t)$ as measured in a frame of reference S are related to the electric and magnetic fields $\mathbf{E}'(x',y',z',t')$ and $\mathbf{B}'(x',y',z',t')$ as measured in a reference frame S' moving with a velocity v_x with respect to S is given by

$$\left.\begin{aligned} E'_x = E_x \quad E'_y = \gamma\left(E_y - v_x B_z\right) \quad E'_z = \gamma\left(E_z + v_x B_y\right) \\ B'_x = B_x \quad B'_y = \gamma\left(B_y + \frac{v_x}{c^2}E_z\right) \quad B'_z = \gamma\left(B_z - \frac{v_x}{c^2}E_y\right) \end{aligned}\right\} \tag{6.1}$$

with, in addition, the usual Lorentz transformation equations for the space time coordinates.

The question then arises as to how the electromagnetic field fits in with the general mathematical formalism presented above. It is first of all clear that the transformation laws given in Eq. (6.1) are not those of a four-vector. For one thing, a four-vector has four components – the electromagnetic field has six, while a second rank tensor has two indices and hence has $4 \times 4 = 16$ components. However, these components need not all be independent. In fact, we can distinguish two important special cases in which the tensor is either symmetric or antisymmetric in its components. In discussing this point, we will work with the contravariant components of a rank 2 tensor as this turns out to be most convenient when dealing with Maxwell's equations in four-vector notation. Thus, we have the two possibilities

$$\left.\begin{aligned} T^{\mu\nu} &= T^{\nu\mu} && \text{symmetric} \\ T^{\mu\nu} &= -T^{\nu\mu} && \text{antisymmetric} \end{aligned}\right\} \tag{6.2}$$

In the symmetric case, we only need to know 10 components, that is $T^{\mu\mu}$, $\mu = 0, 1, 2, 3$ and $T^{01}, T^{02}, T^{03}, T^{12}, T^{13}$ and T^{23}. An important example of a symmetric tensor is the energy-momentum tensor. In the antisymmetric case, we have that

$$T^{\mu\mu} = -T^{\mu\mu} \tag{6.3}$$

so that $T^{\mu\mu} = 0$. What is left are the components $T^{01}, T^{02}, T^{03}, T^{12}, T^{13}$ and T^{23} which automatically give their transposed companions by a change of sign. Written as a matrix, we have

$$T^{\mu\nu} = \begin{pmatrix} 0 & T^{01} & T^{02} & T^{03} \\ -T^{01} & 0 & T^{12} & T^{13} \\ -T^{02} & -T^{12} & 0 & T^{23} \\ -T^{03} & -T^{13} & -T^{23} & 0 \end{pmatrix}. \tag{6.4}$$

Thus only six independent quantities are needed to fully specify the components of an antisymmetric tensor, exactly the same as the number of components of the electromagnetic field. The prospect therefore exists that the electromagnetic field components together constitute the components of a second rank antisymmetric tensor. To test whether or not this is the case, we need to show that an antisymmetric tensor transforms in the same way as the electromagnetic components as given in Eq. (6.1).

The transformation equations for $T^{\mu\nu}$ are given by

$$T^{\mu'\nu'} = \Lambda^{\mu'}_{\mu}\Lambda^{\nu'}_{\nu'}T^{\mu\nu}. \tag{6.5}$$

Expanding the various components we have

$$T^{0'1'} = \Lambda^{0'}_{\mu}\Lambda^{1'}_{\nu}T^{\mu\nu} = \Lambda^{0'}_{0}\Lambda^{1'}_{\nu}T^{0\nu} + \Lambda^{0'}_{1}\Lambda^{1'}_{\nu}T^{1\nu} = \Lambda^{0'}_{0}\Lambda^{1'}_{1}T^{01} + \Lambda^{1'}_{0}\Lambda^{0'}_{1}T^{10} \tag{6.6}$$

where all the other terms vanish either because the diagonal elements $T^{\mu\mu}$ are all zero, or because the elements of the Lorentz transformation matrix are zero. Further, since $T^{01} = -T^{10}$, we get

$$T^{0'1'} = \left(\Lambda^{0'}_{0}\Lambda^{1'}_{1} - \Lambda^{1'}_{0}\Lambda^{0'}_{1}\right)T^{01} = \left(\gamma^2 - \gamma^2\frac{v_x^2}{c^2}\right) = T^{01}. \tag{6.7}$$

In the same way we find that

$$T^{0'2'} = \Lambda^{0'}_{\mu}\Lambda^{2'}_{\nu}T^{\mu\nu} = \Lambda^{0'}_{0}\Lambda^{2'}_{\nu}T^{0\nu} + \Lambda^{1'}_{0}\Lambda^{2'}_{\nu}T^{1\nu} = \Lambda^{0'}_{0}\Lambda^{2'}_{2}T^{02} + \Lambda^{1'}_{0}\Lambda^{2'}_{2}T^{12} \tag{6.8}$$

Substituting for the elements of the Lorentz transformation matrix then gives

$$T^{0'2'} = \gamma\left(T^{02} - \frac{v_x}{c}T^{12}\right). \tag{6.9}$$

Proceeding in this way, we end up with the set of transformation equations:

$$\left.\begin{aligned} &T^{0'1'} = T^{01} \quad T^{0'2'} = \gamma\left(T^{02} - \frac{v_x}{c}T^{12}\right) \quad T^{0'3'} = \gamma\left(T^{03} + \frac{v_x}{c}T^{31}\right) \\ &T^{2'3'} = T^{23} \quad T^{3'1'} = \gamma\left(T^{31} + \frac{v_x}{c}T^{03}\right) \quad T^{1'2'} = \gamma\left(T^{12} - \frac{v_x}{c}T^{02}\right) \end{aligned}\right\} \tag{6.10}$$

which can be compared with Eq. (6.1) written as follows:

$$\left.\begin{aligned} &\frac{E'_x}{c} = \frac{E_x}{c} \quad \frac{E'_y}{c} = \gamma\left(\frac{E_y}{c} - v_xB_z\right) \quad \frac{E'_z}{c} = \gamma\left(\frac{E_z}{c} + v_xB_y\right) \\ &B'_x = B_x \quad B'_y = \gamma\left(B_y + \frac{v_x}{c}\frac{E_z}{c}\right) \quad B'_z = \gamma\left(B_z - \frac{v_x}{c}\frac{E_y}{c}\right). \end{aligned}\right\} \tag{6.11}$$

Thus we can make the identifications:

$$\left.\begin{array}{lll} T^{01} = \dfrac{E_x}{c} & T^{02} = \dfrac{E_y}{c} & T^{03} = \dfrac{E_z}{c} \\ T^{23} = B_x & T^{31} = B_y & T^{12} = B_z. \end{array}\right\} \tag{6.12}$$

Usually the symbol $F^{\mu\nu}$ is used for the components of the electromagnetic field tensor, also known as the Faraday tensor, so we can write:

$$F^{\mu\nu} = \begin{pmatrix} 0 & \frac{E_x}{c} & \frac{E_y}{c} & \frac{E_z}{c} \\ -\frac{E_x}{c} & 0 & B_z & -B_y \\ -\frac{E_y}{c} & -B_z & 0 & B_x \\ -\frac{E_x}{c} & B_y & -B_z & 0 \end{pmatrix} \tag{6.13}$$

The identification made in Eq. (6.12) is not unique since we can also make the identification:

$$\left.\begin{array}{lll} T^{01} = B_x & T^{02} = B_y & T^{03} = B_z \\ T^{23} = -\dfrac{E_x}{c} & T^{31} = -\dfrac{E_y}{c} & T^{12} = -\dfrac{E_z}{c} \end{array}\right\}. \tag{6.14}$$

which leads to the tensor

$$G^{\mu\nu} = \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -\frac{E_z}{c} & \frac{E_y}{c} \\ -B_y & \frac{E_z}{c} & 0 & -\frac{E_x}{c} \\ -B_z & -\frac{E_y}{c} & \frac{E_x}{c} & 0 \end{pmatrix} \tag{6.15}$$

known as the dual tensor.

Thus we have seen that the electromagnetic field can be represented in special relativity by two second rank antisymmetric tensors, the Faraday tensor F with contravariant components given by Eq. (6.13), and its dual G with contravariant components given by Eq. (6.15).

6.2 Dynamics of the Electromagnetic Field

Using the results obtained above, we can show how to rewrite Maxwell's equations in the language of four-vectors. In order to do this, we first of assume that charge is a relativistic scalar, i.e. it is the same in all reference frames. We then introduce a new four-vector, the current density four-vector $\vec{J}$ with contravariant components given by

$$J^{\mu} = \rho_0 u^{\mu}. \tag{6.16}$$

Here the u^{μ} are the contravariant components of the velocity four-vector for which

$$\left.\begin{array}{l} u^0 = \dfrac{c}{\sqrt{1 - u^2/c^2}} \\ u^1 = \dfrac{u_x}{\sqrt{1 - u^2/c^2}} \\ u^2 = \dfrac{u_y}{\sqrt{1 - u^2/c^2}} \\ u^3 = \dfrac{u_z}{\sqrt{1 - u^2/c^2}} \end{array}\right\} \tag{6.17}$$

The current density is evaluated at the point (x, y, z, t) as measured in S, and is determined both by the velocity with which the charges are moving and by the density of charge at this point at this time. However, the charge density ρ_0 is the proper charge density, that is, it is the charge per unit volume as measured in the neighbourhood of the event (x, y, z, t) as measured with respect to a frame of reference that in which the charges at that point are at rest. Thus, in particular, we have

$$J^0 = \rho_0 u^0 = \frac{\rho_0}{\sqrt{1-u^2/c^2}} c = \rho c \tag{6.18}$$

where ρ given by

$$\rho = \frac{\rho_0}{\sqrt{1-u^2/c^2}} \tag{6.19}$$

is the charge density in the frame of reference S in which the length of the volume occupied by the charge has been contracted in the direction of motion of the charge as measured in S. To see what this means, we can suppose that we are considering a small volume ΔV_0 which is stationary with respect to the charges within this volume. But these charges are moving with a velocity $\mathbf{u}$ as measured from a frame of reference S. Thus, if we let $\Delta V_0 = \Delta x_0 \Delta y_0 \Delta z_0$, and the charges are moving in the x direction in S, i.e.$u_y = u_z = 0$, then according to the reference frame S, the x dimension of this volume is contracted to a length

$$\Delta x = \sqrt{1-u^2/c^2}\Delta x_0 \tag{6.20}$$

while the lengths in the other direction are unaffected. Thus, the volume occupied by the charges as measured in S is

$$\Delta V = \Delta x \Delta y \Delta z = \sqrt{1-u^2/c^2}\Delta x_0 \Delta y_0 \Delta z_0 \tag{6.21}$$

so that

$$\Delta V_0 = \frac{\Delta V}{\sqrt{1-u^2/c^2}} \tag{6.22}$$

If we let the charge within this volume be ΔQ, then the charge density will be, in S

$$\rho = \frac{\Delta Q}{\Delta V} = \frac{1}{\sqrt{1-u^2/c^2}} \frac{\Delta Q}{\Delta V_0} = \frac{\rho_0}{\sqrt{1-u^2/c^2}} \tag{6.23}$$

where we have explicitly used the fact that the charge is the same in both frames of reference, i.e. that charge is a relativistic scalar.

One of the important properties of the current density four-vector follows if we calculate the 'four-divergence' of $\vec{J}$:

$$\vec{\partial} \cdot \vec{J} = \partial_\mu J^\mu = \frac{\partial J^0}{\partial x^0} + \frac{\partial J^1}{\partial x^1} + \frac{\partial J^2}{\partial x^2} + \frac{\partial J^3}{\partial x^3} \tag{6.24}$$

$$= \frac{\partial c\rho}{\partial ct} + \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} \tag{6.25}$$

$$= \nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t}. \tag{6.26}$$

The last expression expresses the conservation of charge: the term $\nabla \cdot \mathbf{J}$ is the rate at which charge 'diverges' from a point in space, while the time derivative is the rate of change of the charge density at that point. Since charge is conserved, i.e. neither created or destoyed, the sum of these two terms must be zero, i.e.

$$\partial_\mu J^\mu = \nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0. \tag{6.27}$$

We can now show that Maxwell's equations can now be written in the form

$$\partial_\nu F^{\mu\nu} = \mu_0 J^\mu \tag{6.28}$$
$$\partial_\nu G^{\mu\nu} = 0. \tag{6.29}$$

where μ_0 is the magnetic permeability of free space. To demonstrate this, it is necessary to merely expand the expressions for the four possible values of the free index in each case. For example, we have, on setting $\mu = 0$ in Eq. (6.28)

$$\partial_\nu F^{0\nu} = \mu_0 J^0 = \mu_0 \rho c \tag{6.30}$$

and, on expanding the left hand side:

$$\partial_0 F^{00} + \partial_1 F^{01} + \partial_2 F^{02} + \partial_3 F^{03} = \mu_0 \rho c. \tag{6.31}$$

Replacing the partial derivatives by the usual forms in terms of x, y, and z, and noting that $F^{00} = 0$ gives

$$\frac{1}{c}\frac{\partial E_x}{\partial x} + \frac{1}{c}\frac{\partial E_y}{\partial y} + \frac{1}{c}\frac{\partial E_z}{\partial z} = \mu_0 \rho c. \tag{6.32}$$

Using the fact that $c^2 = (\mu_0 \epsilon_0)^{-1}$ and recognizing that the derivatives on the right hand side merely define the divergence of $\mathbf{E}$, we get

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{6.33}$$

which is Gauss's Law. In a similar way, the other Maxwell's equations can be derived. This is left as an exercise for the reader.