RELATIVITY PHYSICS

Roy E. Turner

RELATIVITY PHYSICS

Roy E. Turner

Reader in Theoretical Physics
University of Sussex

ROUTLEDGE & KEGAN PAUL
London, Boston, Melbourne and Henley

First published in 1984
by Routledge & Kegan Paul plc
39 Store Street, London WC1E 7DD, England
9 Park Street, Boston, Mass. 02108, USA
464 St Kilda Road, Melbourne,
Victoria 3004, Australia and
Broadway House, Newtown Road,
Henley-on-Thames, Oxon RG9 1EN, England

Set in Press Roman by Hope Services, Abingdon
and printed in Great Britain
by Cox & Wyman Ltd., Reading

Library of Congress Cataloging in Publication Data

Turner, R. E. (Roy Edgar), 1931–

Relativity physics.
(Student physics series)
Bibliography: p.
Includes index.
1. Relativity (Physics) I. Title. II. Series.
QC173.55.T87 1984 503.1'1 83-24707

ISBN 0-7102-0001-3

Preface

It is somewhat presumptuous to produce yet another book on relativity, despite the fact that the publishers are producing a new series of texts on undergraduate physics. There is already a number of excellent texts on the market which deal clearly with many aspects of the subject. As always in this situation, the author is reduced to justifying his efforts by claiming that he hopes he has produced a new perspective on the subject. In my case that perspective is not really a new one but goes back to the arguments that were around at the beginning of this century. Most texts at undergraduate level omit, or treat very scantily, the role of electromagnetic theory in the development of special relativity, and yet it played a central role.

I have tried in this book to present mechanics and electromagnetism in the context of special relativity, to explain, in turn, the problems that arose when the principle of relativity was applied to each of them and how Einstein showed that, once the leap in imagination was made, the solution was relatively simple. It is an exciting story and I hope that the book conveys, if only in parts, that excitement. But above all I hope that undergraduates will enjoy reading it, for physics is an enjoyable subject to try to understand.

The reader should have had a good grounding in Newton's equations of motion and an understanding of electromagnetism up to Maxwell's equations and their plane-wave solutions. Quantum mechanics is mentioned but the reader is only required to know the Planck and de Broglie relations. Some readers may find Chapter 7 hard going. The chapter is there for three reasons.

First, it brings together the two subjects that dominate this book, mechanics and electromagnetism. Second, relativistic equations of motion can be written down (despite the subsequent difficulties that they lead to) and very rarely are. Finally I hope it will encourage the interested student to go on and pursue the subject further.

Notation is very important in producing a clear presentation. I have steered away from tensor notation because, elegant though the equations look in this form, I believe that in a first reading it obscures the physics. Ordinary three-vectors are denoted as usual in bold type and their magnitude in normal type. Four-vectors are the same except they are distinguished by an arrow above them. Thus for example, the three-momentum is denoted by **p** and its magnitude by p. The four-momentum or energy-momentum four-vector by $\vec{\mathbf{p}}$ and its 'magnitude' by $\vec{p}$.

Finally I should like to thank my colleagues in the physics division for the many questions that I have put to them over the years that I have given a course on this subject at the University of Sussex. This particularly applies to Drs G. Barton, J. Byrne and J. Plaskett who have tolerated my pestering with good humour. Any mistakes are of course my own.

Contents

Chapter 1

The principle of relativity

1.1 Introduction

There are very few discoveries that can be said, unequivocally, to have changed radically the course of human progress. In evaluating the importance of the subject of this book it is possible to go even further and say that the formulation of special relativity, at the beginning of this century, was the most significant event in the history of mankind. The reasons for this are seen in the simple, and probably the most publicized, scientific formula

$$E = mc^2 \qquad [1.1]$$

which shows the equivalence of energy, E, and mass, m, their values being related by the square of the velocity of light, c. From this apparently innocent-looking result, modern technology has developed the seeds of mankind's own destruction, a fact that was graphically and cruelly illustrated at Hiroshima and Nagasaki. The enormous amount of energy that is released in the atomic bomb is due to the factor c^2, since c is a very large number, $3 \times 10^8 \text{ms}^{-1}$. Thus a mass of 1 kg is equivalent to 9×10^{16} J of energy. (Compare this with the annual output of a 500 MW power station, a typical size for the UK which is equal to $5 \times 10^8 \times (60)^2 \times 24 \times 365 = 1.6 \times 10^{16}$ J.)

Although equation [1.1] is the most quoted result from special relativity, its derivation by Einstein in 1905 was not the most surprising or the most contentious result for the then contemporary physicists. As early as 1881 Sir J. J. Thomson considered a model of the electron which consisted of a uniform charge distributed over the surface of a sphere. The electrostatic energy of the electron is then

$$E = e^2/8\pi\epsilon_0 a$$

where e is the total electron charge and a is the radius of the sphere. He further observed that for a charge e moving with a uniform velocity $\boldsymbol{v}$ the energy of its field is $m_{\text{elm}}\,v^2/2$, where the electromagnetic 'mass' is $m_{\text{elm}} = e^2/6\pi\epsilon_0 c^2 a$. Combining this result with the expression for the energy gives

$$E = \frac{3}{4} m_{\text{elm}} c^2$$

(For a detailed discussion of these results see Feynman, vol. II. For a description of other theories linking energy and mass see Whittaker, vol. 2.)

Another feature of the Einstein theory, namely that the length of an object is not the same when measured by a stationary observer as when measured by a moving observer, was also foreshadowed by Lorentz and FitzGerald. This phenomenon is still known as the Lorentz-FitzGerald contraction.

Despite these precursors, Einstein's theory provoked considerable controversy when it was published. The reasons for this will, I hope, emerge in the course of reading this book. The presentation given here will not be an historical one for three reasons. First, no physical theory proceeds smoothly from one development to the next—many blind and unprofitable alleys are followed. Second, many of the significant historical landmarks are lost in the mists of time. For example, whether Einstein knew of a critical experiment—the Michelson-Morley experiment—is a matter of controversy. (This controversy may have been resolved; see *Physics Today*, vol. 35, no. 8, Aug. 1982.) Thirdly, an excellent and highly readable account is given in Whittaker's classic work *A History of the Theories of Aether and Electricity* (see the bibliography). Despite forgoing the strictly historical approach, however, some of the original chronology will inevitably be followed, not only because this is the logical description but also because, hopefully, some of the original excitement will be conveyed.

As we shall see, special relativity is not a theory in the same sense as solid state physics or electromagnetism. What relativity

—or, more precisely, the principle of relativity—is, is a kind of 'check' on the validity of other theories. Crudely speaking the principle of relativity says that the laws of physics are the same for a stationary observer as they are for an observer moving with a constant velocity with respect to the first. Thus we can apply the principle of relativity to any physical theory and see if it is satisfied. If it is not, then we assert that the theory cannot be wholly correct but, at the best, can only be an approximation to a correct, relativistically invariant theory. (The precise meaning of the phrase 'relativistically invariant' will be given later. Here it may be taken to mean 'satisfies the principle of relativity'.) In fact for over two centuries it was believed that Newton's equations gave a correct description of the motion of material bodies. Following Einstein's discoveries, however, they were found not to satisfy the principle of relativity. Given their experimental verification, the conclusion to be drawn (and this also followed from the mathematics) was that in some sense Newton's equations were only valid as an approximation. The approximation was that the velocities of all the particles should be small compared with the velocity of light, a condition well satisfied by the motion of the planets. The Earth's velocity around the Sun, for example, is approximately $3 \times 10^4\,\mathrm{ms}^{-1}$ and the velocity of light is $3 \times 10^8\,\mathrm{ms}^{-1}$, so Newton's solution for the motion of the Earth should be accurate to one part in 10^4. (In fact it is the square of the velocity ratio which is important, so that the limit of validity is even better—one part in 10^8.)

Finally in this introduction a comment on the use of the word 'relativistic'. We are using it combined with the word 'invariant' to mean a theory which satisfies the principle of relativity. It is common these days, however, to refer to 'relativistic expressions' for some physical quantity, e.g. momentum. What this really means is that the expression so referred to is consistent with the principle of relativity. This is a little misleading, since Newton's expression for momentum, for example, is approximately compatible (see the discussion above) with the principle of relativity but is not described as relativistic. Probably the more correct way to describe the expressions derived by Einstein and others is 'fully relativistic', as opposed to the Newtonian expressions which are only 'approximately relativistic'. We shall try to avoid the adjective 'relativistic'.

This book is about a number of branches of physics (mainly two) in which the correct application of the principle of relativity (as realized by Einstein) radically altered their development. It may thus be reasonably labelled 'Relativity Physics'.

1.2 Reference systems

Physics is about measurement and the relationship of one measurement to another. Even in the most primitive societies there are two concepts which are intuitively obvious and present no difficulty in understanding. These two concepts are space and time. Furthermore, these two concepts are uncluttered, not being endowed with mystical or magical qualities. Very early on in his development man put these concepts on a quantitative basis. He used the rotation of the Earth as his clock and introduced an arbitrary length against which all other lengths were to be compared. For each of these measurements he chose a unit. For time the unit eventually became the second, defined until relatively recently (the modern definition will be given later) as $(1/24) \times (1/60) \times (1/60)$ of the period of rotation of the Earth. For distance the arbitrary length has varied from culture to culture, e.g. the idyllic-sounding rod, pole or perch and the rather dull-sounding metre.

If we are concerned with measuring distance—for example, the distance travelled by a particle—then this is most easily done by setting up a coordinate system or reference frame. The simplest such reference frame is a Cartesian one, which consists of three axes at right angles, labelled x, y and z. The position of any event is then given by specifying the coordinates (x,y,z), i.e. the distance one has to move along the x, y and z axes to the point in space where the event occurred. Figure 1.1 shows the trajectory of a particle and (x_1, y_1, z_1) and (x_2, y_2, z_2) are the coordinates of two events, namely the particle passing two markers.

For the measurement of time we need a clock which, with reference to some arbitrary zero, will specify how many seconds have elapsed before a particular event occurs. The instrument we

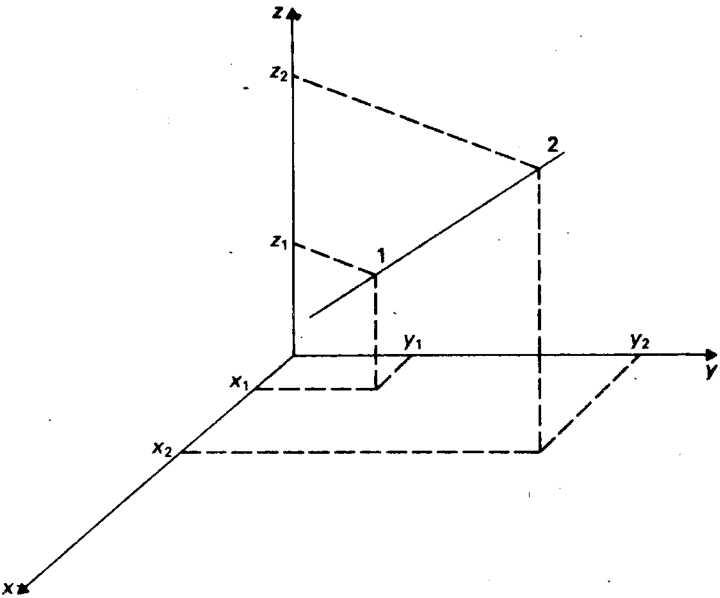

Fig. 1.1 The trajectory (linear) of a particle that passes two markers with coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2).

use as a clock will be discussed in Chapter 3, but for the moment a laboratory stopclock will suffice. Thus for the two events illustrated in Fig. 1.1 we can add a fourth 'coordinate', time. The particle passes the first marker at a position in space given by (x_1, y_1, z_1) and does so at time t_1. We say it has the space-time coordinates (x_1, y_1, z_1, t_1). Similarly for event 2 we have the space-time coordinates (x_2, y_2, z_2, t_2). Thus to specify the coordinates of any event completely we need to know the three spatial components and the time; if we use a vector notation, this may be shortened to $(\mathbf{r}, t)$, where

$$\mathbf{r} = \hat{\mathbf{i}}x + \hat{\mathbf{j}}y + \hat{\mathbf{k}}z \qquad [1.2]$$

is the position vector.

Now the absolute values of $(\mathbf{r}, t)$ cannot be important because where we have the origin of the coordinate system is quite arbitrary, as indeed was the choice of the zero of time. In fact we are only interested in relative positions. Thus we might wish to know the time between the two events depicted in Fig. 1.1, i.e.

$$\Delta t = t_2 - t_1$$

This is independent of the origin of time. This is because if we shift the origin of time by an arbitrary amount, say t_0, then the new absolute times are $t_1' = t_1 - t_0$ and $t_2' = t_2 - t_0$, but the difference between the two, as we can see, remains the same. Similarly the distance travelled by the particle is independent of the origin of the reference frame. From Fig. 1.1 it is clear by simple geometry that

$$l^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 \qquad [1.3]$$

where l is the distance travelled by the particle. Since this expression depends only upon the differences of the coordinates, by arguments similar to those used for time a shift in the origin of all or some of the spatial coordinates will not change the value of l.

1.3 The homogeneity and the isotropy of space

Δt and l are examples of invariants. Formally, Δt is invariant with respect to translations of the origin of time, whilst l is invariant with respect to translations of the spatial origin. This arbitrariness in the origin of both space and time is sometimes referred to as the homogeneity of space and time respectively.

That the length, l, is the same when measured in a coordinate system which is rotated with respect to the first is easily proved (see problem 1, Chapter 1). This leads to the conclusion that there is no preferred direction in space. Space is also said, therefore, to be isotropic. Actually we have really inverted the argument; the physical assumption is that space is both homogeneous and isotropic. Hence we can use Euclidean geometry to show that length is an invariant. Similarly it is the physical assumption of the homogeneity of time which leads to the time interval being an invariant. However, the correctness of Euclidean geometry comes more naturally to most people than the more abstract concepts of homogeneity and isotropy of space.

The above discussion has gone into some detail to emphasize that reference frames, on the assumptions of the homogeneity of space and time and the isotropy of space, are arbitrary. It follows therefore that any physical law must not depend upon the particular reference frame we have chosen to select. To give an example, the equations of motion of two particles in one dimension acting through a force $F(x_1, x_2)$ are

$$
\begin{aligned}
m_1 \frac{d^2 x_1}{dt^2} &= F(x_1, x_2) \\
m_2 \frac{d^2 x_2}{dt^2} &= -F(x_1, x_2)
\end{aligned}
\qquad [1.4]
$$

where x_1 and x_2 are the coordinates of particles 1 and 2 in a particular reference frame and m_1 and m_2 are their masses. Now suppose we describe the same two particles by using a reference frame whose origin is at x_0 in the original frame. Then if we call the coordinates of the particles in the new frame x_1' and x_2', then

$$x_1 = x_1' + x_0 \text{ and } x_2 = x_2' + x_0$$

Since x_0 is a constant, its derivative with respect to time is zero and hence if we substitute these transformation equations into the equations of motion (equations [1.4]) we get

$$
\begin{aligned}
m_1 \frac{d^2 x_1'}{dt^2} &= F(x_1' + x_0, x_2' + x_0) \\
m_2 \frac{d^2 x_2'}{dt^2} &= -F(x_1' + x_0, x_2' + x_0)
\end{aligned}
$$

the equations of motion in the new reference frame.

Now there are two things wrong with this result.

1. The new equations depend upon the origin of the coordinate system, x_0, which clearly offends against the homogeneity of space.
2. The force of interaction does not have the form $F(x_1', x_2')$, which it should if the equations are to have the same form as in the original reference frame.

This latter point means that these equations of motion do not satisfy the principle of relativity.

Now, earlier on we discussed the fact that, in going from one frame to another with a different origin, the invariant was not the absolute position of one event but the distance between two events. This suggests that if the forces between two particles always depended only upon the distance between the particles then we would remove the difficulties discussed above. In fact we find in nature that all forces between particles have this property. (Strictly speaking, although the magnitude of forces in nature depends upon their distance apart, their direction is not always along the line of centres. These are known as non-central forces.) Thus we can write

$$F(x_1, x_2) = f(x_1 - x_2)$$

and hence in the new reference frame the equations of motion become

$$\begin{aligned} m_1 \frac{d^2 x_1'}{dt^2} &= f(x_1' - x_2') \\ m_2 \frac{d^2 x_2'}{dt^2} &= -f(x_1' - x_2') \end{aligned} \qquad [1.5]$$

These equations are now independent of x_0 and have the same form as the original equations.

The requirement that the equations of motion look the same in all reference frames is known as requiring that they be form invariant. Ensuring that this was so in the above example was in fact applying the principle of relativity. However, before explicitly giving and explaining this principle, there is one other transformation between coordinate systems that we should investigate.

1.4 Reference frames with a constant relative velocity

We have already agreed that space is homogeneous (as well as isotropic), i.e. if we displace the origin of the reference frame it should make no difference to any physical measurement or law. What happens, however, if we have a frame that is moving with respect to an original frame? Should length be an invariant and should the laws of physics be form invariant?

Before examining these questions it will be helpful to derive the transformation from the original reference frame to the moving frame. For so-called 'special relativity' we need only concern ourselves with the case when the relative velocity of the frames is constant. In particular, let the respective y and z axes of the two frames be parallel (because of the assumed isotropy of space we can always rotate one of the sets of axes to ensure that this is so). Further, let the x' axis of the second frame be moving along the x axis of the first frame with a constant velocity, v. We shall refer to the original frame as the K frame and the moving frame as the K' frame. Alternatively, and somewhat more loosely, we shall refer to the laboratory frame and the moving frame. The corresponding situation in two dimensions is illustrated in Fig. 1.2.

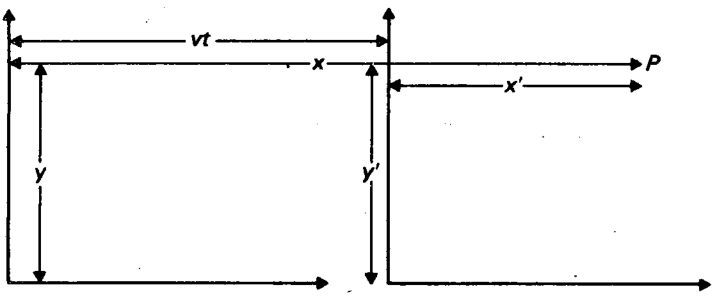

Fig. 1.2 The K' frame is travelling with a velocity along the x axis of the K frame. The y and y' axes are parallel. It is assumed that at $t = 0$, the origins are coincident and subsequently that time is the same in both frames, i.e. $t' = t$ for all t. The point P has the coordinates x, y in the K frame and x', y' in the K' frame. Geometry clearly shows that $y' = y$ and $x' = x - vt$.

The origins are coincident at $t = 0$ and we make the further assumption that time is the same in K as K'. The equations of the transformation, known as the Galilean transformation, are, from Fig. 1.2,

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad [1.6]$$

The last of these equations comes from the analogy with $y' = y$, i.e. it is an axis at right angles to the direction of motion. It follows immediately, by differentiating with respect to time, that the equations for transforming the velocities in the K frame to their corresponding velocities in the K' frame are

$$u_x' = u_x - v \qquad u_y' = u_y \qquad u_z' = u_z \tag{1.7}$$

A second differentiation shows that the acceleration in the two frames is equal (remember that v is a constant), i.e.

$$a_x' = a_x \qquad a_y' = a_y \qquad a_z' = a_z \tag{1.8}$$

Now the question is: Can we distinguish between these two frames? We have already assumed that space is homogeneous and, given this, it would seem that the frames K and K' are indistinguishable. All we can say is that they have a relative velocity v. This is equivalent to making a small displacement dx in a time dt. The K' frame moving with a constant velocity v corresponds to a series of displacements dx each in time dt, such that dx = vdt. From this point of view the equivalence of K and K' is expressing nothing more than the homogeneity of space. However, this is not the whole story.

Suppose that in some sense space is absolute, i.e. it exists in its own right, or, to put the proposition another way, an object occupies a particular portion of space and if it is moved it no longer occupies that piece of space but another and different portion. Newton believed in such an absolute space and devised several ingenious experiments to demonstrate its existence. Other philosophers such as Leibniz and Mach denied the existence of absolute space and preferred to think of space as a series of relationships between objects. For the moment let us take Newton's side in the debate. Then it is clear that frames moving with a relative velocity can be distinguished, for there exists a reference frame in which absolute space is at rest. All other frames can be distinguished by their velocity with respect to the absolute frame. So if we have two frames moving along the x axis of the absolute frame with velocities v_1 and $v_1 + v$, then their relative velocity is v. Nevertheless, they can be distinguished because one has a velocity v_1 and the other $v_1 + v$ with respect to absolute space.

On the other hand, if we follow Leibniz and Mach and dispense with absolute space, then all we can say is that the two frames have a relative velocity, and whether one is moving and the other