SPECIAL RELATIVITY

THE FOUNDATION OF MACROSCOPIC PHYSICS

W. G. DIXON

SPECIAL RELATIVITY

THE FOUNDATION OF MACROSCOPIC PHYSICS

W. G. DIXON
Fellow of Churchill College, Cambridge

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON · NEW YORK · MELBOURNE

Published by the Syndics of the Cambridge University Press
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
Bentley House, 200 Euston Road, London NW1 2DB
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1978

Printed in Great Britain
at the University Press, Cambridge

Library of Congress Cataloguing in Publication Data

Dixon, William Graham, 1940–
Special relativity

Includes bibliographical references and index

1. Relativity (Physics) I. Title

QC173.65.D58 530.1'1 77–83991
ISBN 0 521 21871 3

Preface

The special theory of relativity is often considered as irrelevant to the macroscopic physics of ordinary material systems. The range of velocities, pressures and temperatures encountered under terrestrial conditions is such that the differences between the Newtonian and relativistic theories are negligibly small. Either theory can thus be used, and as the Newtonian theory is usually considered to be the simpler, it is the one usually adopted. But *is* the Newtonian theory actually the simpler of the two? This depends on what one is trying to do. Ballistic calculations are undoubtedly made more complicated by the use of relativistic formulae in place of the corresponding Newtonian ones, but special relativity has more to offer than nuisance factors of $\sqrt{(1-v^2/c^2)}$. The aim of this book is to show that an understanding of the basic laws of macroscopic systems can be gained more easily within relativistic physics than within Newtonian physics. The speed of the systems concerned is irrelevant. Even equilibrium thermodynamics gains by being seen from a relativistic viewpoint.

The book is not directed towards any particular university course. It tries to show the unity of dynamics, thermodynamics and electromagnetism under the umbrella of special relativity, and it should be accessible to any second year undergraduate in mathematics or physics. The emphasis throughout is on the extraction by systematic development of a maximum of information from a minimum of assumptions. With this in mind, the first chapter lays minimal physical foundations for the special theory of relativity and explores its relationship to Newtonian physics. The assumption that the speed of light is independent of the motion of the observer is found to be unnecessary. A prior knowledge of special relativity is not essential but an acquaintance with its basic ideas will be helpful. The second chapter lays the mathematical foundations needed for the subsequent development. The remaining three chapters develop the foundations of particle and continuum dynamics, and the thermodynamics and electrodynamics of fluids, within this relativistic framework. Fluids are studied in preference to solids as they are conceptually simpler. As four-dimensional spacetime techniques are used throughout, much of the theory can be taken over into general relativity with little alteration.

The S.I. system of units that is now widely adopted for the presentation of formulae in electromagnetism does not combine naturally with the four-dimensional tensor formulation that is used in special relativity. For this reason Gaussian (c.g.s.) units have been used instead for the development of electrodynamics in Chapter 5.

Chapters are divided into sections, and equations are numbered consecutively within each section. These numbers run continuously through the subsections into which some sections are divided, thus §4 of Chapter 5 has subsections labelled 4*a* to 4*c* and equations numbered (4.1) to (4.42). Sections and equations within the current chapter are referred to simply by these numbers. References to sections and equations of another chapter are prefixed by the chapter number and a hyphen, thus §4*a* means subsection *a* of section 4 of the current chapter but §4-1 means section 1 of Chapter 4. Reference to publications is by author and year. Details of these publications are given in a list at the end of the volume.

This book was begun during leave of absence from Churchill College, Cambridge, for the academic year 1974/75. I am grateful to Churchill College for financial support during that period, and to the Department of Physics and Astronomy, University College London, for its hospitality during it. I would also like to thank my wife for her patience and constant encouragement during my writing and typing of the book.

March 1978 W. G. Dixon

Contents

Preface page ix

1 The physics of space and time

1 Introduction 1
2 Frames of reference 2
3 Newtonian conceptions 8
4 Foundations of the special theory of relativity 11
5 The transformation between inertial frames 15
6 The uniformity of space and time 21
7 The invariant measure of interval 25
8 Something for nothing? 30
9 The Principle of Equivalence 32
10 Gravitation as curvature 35

2 Affine spaces in mathematics and physics

1 Introduction 42
2 Basic properties of affine spaces 44
2*a* Algebraic properties 44
2*b* Geometric properties 51
3 Fundamental tensors 55
4 Specializations and limits 62
4*a* Cartesian tensors 62
4*b* The Newtonian limit $c \to \infty$ 64
5 Symmetries and alternating tensors 68
5*a* The bracket notations 68
5*b* Oriented affine spaces 72
6 Integration and orientation for surfaces in E_n 76
6*a* Definitions 76
6*b* Stokes' Theorem 80
7 Spacetime geometry in special relativity 86
8 Spacetime geometry in Newtonian physics 92
8*a* Intrinsic theory 92
8*b* The Newtonian situation as a limiting case 96

3 Foundations of dynamics

1 The free motion of particle systems *page* 99
1*a* Conservation laws 99
1*b* Centroid lines 104
2 Mass and energy in Newtonian physics 108
2*a* Three-dimensional formulation 108
2*b* Four-dimensional formulation 111
3 Mass and energy in the relativistic theory 113
3*a* The relativistic unification of mass and energy 113
3*b* The relativistic distinction between mass and energy 115
3*c* The non-relativistic limit 118
4 The free motion of a continuum 121
4*a* Consequences of momentum conservation 121
4*b* Consequences of angular momentum conservation 124
4*c* The symmetric energy–momentum tensor 126
5 Spacetime decomposition of $T^{\alpha\beta}$ 128
6 Three-dimensional interpretation of $T^{\alpha\beta}$ 135
7 Energy in Newtonian continuum mechanics 140
7*a* The Newtonian stress–energy tensor 140
7*b* Relation to the relativistic description 143

4 Relativistic simple fluids

1 The role of thermodynamics in continuum mechanics 146
2 The entropy law 148
3 The nature of equilibrium 151
4 Thermostatics of a relativistic simple fluid 152
4*a* Detailed characterization of a fluid in equilibrium 152
4*b* The local state in equilibrium 155
4*c* The global state in equilibrium 159
4*d* Hyperbolic motion 161
5 Description of near-equilibrium states 163
6 Phenomenological laws – first approximation 166
6*a* The entropy source strength near local equilibrium 166
6*b* The Lie derivative – an aid to interpretation 168
6*c* General form of the phenomenological laws 170
6*d* Implications of the entropy law 172
6*e* Linearization 174

7 Phenomenological laws – second approximation *page* 176
7*a* Forces, fluxes, impedance and relaxation 176
7*b* The rate-of-change operator $D/D\tau$ 177
7*c* The entropy source strength and its implications 179
8 Thermal diffusion 182
9 The Newtonian limit 183
10 Perfect fluids 187

5 Electrodynamics of polarizable fluids

1 The role of thermodynamics in electrodynamics 189
2 Maxwell's equations in spacetime form 191
2*a* The electromagnetic field tensor in vacuo 191
2*b* The charge–current vector 192
2*c* Polarization and magnetization 195
2*d* The field tensors in a polarizable medium 199
3 The electromagnetic energy–momentum tensor in vacuo 202
4 Local equilibrium of a polarizable simple fluid 204
4*a* Characterization of the fluid and its equilibrium 204
4*b* Implications of Lorentz invariance 206
4*c* Provisional Gibbs relations 208
5 Interpretation of the conditions for global equilibrium 210
5*a* The stationary nature of the equilibrium state 210
5*b* The parameter λ as a universal constant 211
6 The Kelvin variables 214
6*a* The Gibbs relations reconsidered 214
6*b* Global equilibrium reconsidered 218
6*c* Motion of a perfect polarizable fluid 221
7 Linear constitutive relations 222
7*a* Electric and magnetic susceptibilities 222
7*b* The Helmholtz variables 223
7*c* The energy–momentum tensors of Abraham and Minkowski 225
8 Near-equilibrium states 228
8*a* The comparison equilibrium state 228
8*b* The entropy source strength – first approximation 229
8*c* The entropy source strength – second approximation 232

9 Linearized phenomenological laws *page* 235
9*a* The general situation 235
9*b* The weak-field case 237
9*c* Electromagnetic dispersion 240

Appendix. Vector and dyadic notation in three dimensions 243

Publications referred to in the text 245

Summary and index of symbols and conventions 247

Subject index 253

1
The physics of space and time

1 Introduction

The special theory of relativity has its historical origin in a study of electromagnetic phenomena. It takes its name from its denial of the concept of absolute motion and the consequent recognition that only relative motion has any physical significance. However, it does recognize a preferred class of observers who are in uniform motion relative to one another, even though it denies that it is meaningful to ask which of them is at rest in any absolute sense. Hence the qualification 'special', the hope being that it would ultimately be superseded by a theory in which all observers are treated as equivalent.

At the time that the special theory was being developed, around the beginning of this century, it was believed that all forces in nature would ultimately be reducible to electromagnetism and gravitation. With the success of the special theory in resolving the conflicts that had existed between Newtonian dynamics and Maxwell's electromagnetic theory, it became natural to try to fit gravitation into this new physical framework. That this proved so difficult seems perhaps more surprising now than it did at that time. It is now realized that the ultimate structure of matter is considerably more complicated than was suspected seventy years ago, when the quantum theory was still in its infancy and even the Bohr theory of the atom was still in the future. Although the forces that occur within the atomic nucleus are not yet fully understood, tremendous progress has been made, and underlying it all is the basic framework provided by the special theory of relativity. This is indeed the main strength of the theory. The fact that it predicts modifications of Newtonian dynamics for particles whose speeds are comparable with that of light is important, but its real achievement has been in providing a foundation on which almost the whole of modern physical theory has been built. However, this increasing scope of the special theory has also seemed to increase the apparent perversity of gravitation in refusing to be fitted into this growing structure.

A study of the foundations of the special theory should reveal the origins of its limitations as well as of its successes. Gravitation must thus be expected to play a distinctive part in such a study, inasmuch as the reason for its exclusion from the theory should become clear. But although it is excluded from the theory, it cannot be excluded from the laboratories in which terrestrial physical experiments are performed. To understand the validity of the special theory in such circumstances, some knowledge is required of the modifications which are required to allow for the presence of gravitation. These modifications form the basis of the *general theory of relativity*, so named because Einstein considered that these same modifications also place all observers on an equal footing.

The programme of the present chapter is to give a physical basis for the mathematical models of space and time used in relativity theory. For the reason given above, both the special and general theories will be considered. The mathematical and physical developments of the subsequent chapters will however be confined to the special theory. The physical results used will be ones which hold also in Newtonian theory. Consequently no details will be given of the experimental evidence in their support – the success of the Newtonian theory over a wide range of conditions is sufficient evidence in itself. When such results are particularly simple and have far-reaching implications, they may be dignified with the description 'principle'. This is not intended as a claim that they are 'obvious', but instead that they are firmly supported by the success of Newtonian theory. It will not be necessary to assume the constancy of the speed of light. This speed, 'c', is a fundamental constant of nature which is not primarily connected with electromagnetic phenomena, and a development based on properties of light gives electromagnetism an unnecessary prominence.

2 Frames of reference

It is quite impossible to make any physical statement at all without some implicit assumptions about the nature of space and time. The best that can be done in an investigation of these fundamental concepts is to try to be as explicit as possible about the assumptions that are being made. Our first task must thus be to provide ourselves with a language with which we can discuss the physical world, and which is as free as possible from undefined terms. This involves setting up frames of reference, and hence requires an examination of the concept of such a frame.

Until the advent of the theory of relativity, space and time were believed to be independent and absolute. In the historical development of special relativity, absolute time was the first of these to fall. By assuming the constancy of the speed of light, and examining the practical process of synchronizing clocks using light rays, Einstein showed in 1905 that simultaneity is not an absolute concept – it depends on the motion of the observer. The concept of an absolute space with a fixed three-dimensional Euclidean geometry survived for a further three years, although it was necessary to ascribe rather peculiar behaviour to rods and clocks in motion in order to retain it. But in 1908, Minkowski (1908*a*) showed that the natural framework within which to express special relativity is to consider space and time united to form a single four-dimensional continuum. To quote him in translation: 'Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.'

This union will be taken as our starting point. It may not seem much of an assumption, as no particular geometry is yet being ascribed to this spacetime continuum. So it is worth a pause to consider just what are the assertions about the physical world that are hidden within it. There are essentially two. The first is that space and time are continuous, which may be questioned in the light of the quantum nature of so much of physics. The second is that if two events appear coincident in both space and time to one observer, then they appear so to every other observer. It is difficult to envisage the implications of this being false, but it is not logically impossible. No attempt will be made here to justify these assumptions, but it is good to make clear that they are there as the basis of our subsequent development.

Since our everyday language and experience is based on a separate space and time, it is necessary to begin by considering in general terms how this separation is made by a scientific observer. The qualitative nature of this separation as perceived by our senses will be accepted without analysis. But a scientific observer must make this quantitative. His first step must be to make a clock. It is convenient to envisage this as a hand turning continuously around a graduated dial, together with a counter to count complete revolutions. The hand should turn smoothly (an intuitive concept based on the assumed continuity of time), but until some dynamics has been put into our theory, we cannot ask that it should turn uniformly as this is a concept that needs further definition. With this clock he can 'time' events in his immediate locality, but before he can 'time' distant events, he must decide on an operational definition of simultaneity for widely separated events.

Having done so, he can unambiguously say 'when' any event occurs, but not 'where' it occurs. To do this, he needs also to decide what is meant by the same point of space at different times, i.e. he needs a standard of 'rest'. The 'where' of an event then becomes meaningful, but for him to be able to communicate this information to anyone else, he must also set up a spatial coordinate system. Again, such coordinates are naturally required to vary smoothly from place to place (also supposed intuitive), but apart from this, all that can be said is that three coordinates will be required to specify a location uniquely. (It is perhaps worth noting that one could get away with only a single spatial coordinate if the smoothness requirement were dropped, e.g. by interleaving the decimal expressions of three smooth coordinates so that (0·114, 0·225, 0·336) becomes 0·123123456, but since physical measurements cannot be made with infinite precision, non-smooth coordinate systems are useless for physical purposes.) When this has been done, he will have set up a complete coordinate system for the spacetime continuum which enables every event to be uniquely specified by four coordinates, three being spacelike and one timelike.

To clarify the procedure, we give an example of a way in which these various constructions may be made. It is not intended, however, to be any more fundamental than any other method. This is the radar method. Suppose the observer sends out a pulse of light at time t_1, which is reflected by a distant object and arrives back at the observer at time t_2. Then the instant of reflection is allocated a time coordinate $\tau = \frac{1}{2}(t_1+t_2)$ and a radial distance $r = \frac{1}{2}(t_2-t_1)$. The direction of the reflected pulse may be specified by two angular coordinates (θ, ϕ) which together with r make up the three spatial coordinates of the object at the instant of reflection. The rest state is then characterized by the constancy of the spatial coordinates. If one wishes to envisage the measurement of the angular coordinates, one can think of the observer as being surrounded by a transparent sphere with a grid of latitude and longitude lines marked on it.

In this example the coordinate system is constructed first and the definitions of simultaneity and of rest then follow in the obvious way.

Although this is likely to happen in practice, it is conceptually preferable to think of simultaneity and rest being defined before the coordinates are constructed. For these are clearly physical concepts, while coordinates belong to mathematics. It will be useful to try to keep track of what belongs to mathematics and what to physics in the initial development of the theory, and for this purpose a distinction will be drawn between the physical concept of a frame of reference and the mathematical one of a coordinate system. This will be abstracted from common usage, which makes such a distinction even though it is seldom made explicit.

The coordinate system concept is simple, although coordinates will be allowed which are more general than those used in elementary physics. All that is essential in a coordinate system for spacetime is that there should be four coordinates, each of which varies smoothly, and independently of the other three. This degree of generality is necessary for the time being as so little has so far been assumed about the physical world. The preferred coordinate systems usually used in special relativity and in Newtonian physics can only be introduced after further physical assumptions have been made. But more of this later.

One other feature of these generalized coordinate systems is that it will not necessarily be assumed that the whole of spactime is covered by a single nondegenerate coordinate system. Sometimes it is simply convenient to use coordinate systems with degenerate points, e.g. plane polar coordinates (r, θ), where the origin is degenerate as θ is indeterminate there. In this case degeneracies could be avoided by the use of Cartesian instead of polar coordinates. But in other circumstances one may have no choice in the matter. On the surface of a sphere, for example, there is no coordinate system which covers the whole surface without degeneracy. Hence, to avoid any implicit assumptions about the global topological structure of spacetime, coordinate systems will be allowed which cover only a portion of spacetime. If one considers operational definitions of coordinate systems such as the radar method described above, in a finite time it is possible to survey only a finite volume of space, and so such coordinates are naturally restricted in this way.

If, in this same example, the observer decided to transform from the polar type of coordinate system that he has constructed by direct measurement to a rectangular type of coordinate system by the mathematical transformation $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$, this would not usually be considered as a change of reference frame. But if he set his transparent sphere, with its angular grid, in rotation (relative to its initial state – absolute rotation has not yet been defined), one would say that he was then using a frame of reference that was rotating relative to the initial one. Viewed as coordinate transformations, the difference between these two cases is that in the latter case the transformation of the spatial coordinates is time-dependent, while in the former case it is not. If this is taken as a characterization of those coordinate transformations which are not regarded as changing the corresponding reference frame, then what is left as belonging specifically to the reference frame is just the observer together with the definitions of rest and simultaneity.

This enables us to talk meaningfully about space and time separately in a given reference frame, but that is about all it does allow. Too much has been removed, and what is left is of little use. It would be preferable to leave some structure in the reference frame which is unaffected by a change from rectangular to polar coordinates, but which, say, makes it meaningful to talk about uniform motion in a straight line. This may be achieved by giving a suitable geometric structure to space and, more trivially, also to time. This does not involve any new assumption about the physical world, as no 'reality' will be attributed to the geometry. It is just a step in the construction of a language with which to discuss physical phenomena. One possible geometric structure for time is provided by any arbitrarily constructed clock, or equivalently by the time coordinate of any spacetime coordinate system. Two time intervals are simply *defined* to be equal when such a clock measures them as equal. The clock also gives a unit of time, which when taken together with this definition of equality of interval gives time the metric structure of the real line.

An equally arbitrary construction will be used for the spatial geometry. If a suitable spacetime coordinate system is given, each of the three-dimensional spaces of constant time can be considered as having that three-dimensional Euclidean geometry in which the given spatial coordinates are rectangular Cartesian. These coordinates also provide a unit of length for this geometry. The description 'suitable' is intended to allow for the possibility that some coordinate systems may be better interpreted as, say, spherical polar than rectangular Cartesian. To include these in the procedure, they should first be transformed to a corresponding rectangular system before the geometry is abstracted.

The frame of reference associated with the coordinate system will be taken as consisting of (*a*) the definitions of rest and of simultaneity which are used to separate space and time, (*b*) the corresponding metric structure for time, which includes a unit of time, and (*c*) the corresponding three-dimensional Euclidean geometry for space, together with its unit of length. For conformity with our allowing coordinate systems covering only a portion of spacetime, frames of reference must similarly be allowed in which these geometric structures also only cover portions of space and time. The initial step discussed above is now complete, for all the language permitted by this rich structure can now be used unambiguously, while the underlying assumptions as to the nature of space and time have been made explicit.

The above construction, which proceeds from a coordinate system to a reference frame, raises the question of the extent to which such a frame determines the coordinate system from which its geometric structure was abstracted. Let us say that (x, y, z, t) are *natural* coordinates for a frame if:

(i) Simultaneity of two events corresponds to equality of t,

(ii) The state of rest corresponds to constancy of (x, y, z),

(iii) (x, y, z) are rectangular Cartesian coordinates in space, which agree with the length unit of the frame, and

(iv) t measures time consistently with the metric structure given by the frame.

Then the original coordinate system is a natural one, but the frame alone does not distinguish it from any other natural coordinate system. It is convenient to write the coordinates (x, y, z) as a (3×1) column vector $\mathbf{x}$. We see that a second coordinate system $(\mathbf{x}^*, t^*)$ is natural for some frame if and only if it is related to a given natural coordinate system $(\mathbf{x}, t)$ for that frame by a transformation of the form

$$\mathbf{x}^* = A\mathbf{x} + \mathbf{a}, \quad t^* = t + k, \tag{2.1}$$

where A is a (3×3) orthogonal matrix, $\mathbf{a}$ is a (3×1) column vector and k is a scalar. For future reference it should be emphasized that A, $\mathbf{a}$ and k are constant, and thus in particular independent of time. Geometrically, $\mathbf{a}$ and k represent a change of origin in both space and time, while A describes a rotation or reflection of the spatial coordinate axes.

3 Newtonian conceptions

Let us now consider in more detail the assumptions underlying Newtonian dynamics, as only by so doing can we fully appreciate the origins of the special theory of relativity. As has already been remarked, underlying all Newtonian thought is the concept of absolute time. This comprises more than a belief in the meaning of absolute simultaneity for spatially separated events. It also implies a metric structure for time, so that the equality of two time intervals is also a primitive undefined concept. Once absolute simultaneity is assumed, 'space' becomes absolute in the sense of being the same for all observers. Based on the idealization of the perfectly rigid rod as a measure of distance, Newtonian physics also implicitly assumes that the geometry of space as surveyed with such rods is Euclidean.

For the time being, let us not question these assumptions, but instead investigate the dynamical laws based on them. It is then possible to restrict attention to those frames of reference which are compatible with these natural geometries for both space and time, and with fixed but arbitrary units of both length and time. Such frames will be said to be *allowable*. It is easily seen that the relation between the natural coordinates of any two such allowable frames must have the form

$$\mathbf{x}^* = A(t)\,\mathbf{x} + \mathbf{a}(t), \quad t^* = t + k, \tag{3.1}$$

where again k is constant, but now the orthogonal (3×3) matrix A and the (3×1) column vector $\mathbf{a}$ may be time-dependent. Conversely, if two frames have natural coordinates which are so related, and if one of them is allowable, then so is the other.

It is here that the explicit development of Newtonian dynamics starts. Its first step is to pick out a subset of the allowable frames which are dynamically privileged, by means of the *Principle of Inertia*, otherwise known as Newton's First Law of Motion. This states that:
There exists a family of reference frames in which any particle will continue in its state of rest or of uniform motion in a straight line unless it be compelled by some external force to change that state.
Such frames are said to be *inertial*.

Suppose now that (3.1) connects two inertial frames. Then the state of uniform motion $\mathbf{x}(t) = \mathbf{u}t + \mathbf{c}$ must correspond to uniform motion in the second frame for all values of $\mathbf{u}$ and $\mathbf{c}$. But it follows from (3.1) that

$$\frac{d^2\mathbf{x}^*}{dt^{*2}} = \frac{d^2A}{dt^2}(\mathbf{u}t + \mathbf{c}) + 2\frac{dA}{dt}\mathbf{u} + \frac{d^2\mathbf{a}}{dt^2},$$