DAMAGE BOOK

UNIVERSAL
LIBRARY
OU_158925
UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 530.1 E21R Accession No. 19018

Author Eddington Sir Arthur

Title Relativity theory of Protons & electrons.

This book should be returned on or before the date last marked below.

RELATIVITY THEORY
OF
PROTONS AND ELECTRONS

LONDON
Cambridge University Press
FETTER LANE

NEW YORK • TORONTO
BOMBAY • CALCUTTA • MADRAS
Macmillan

TOKYO
Maruzen Company Ltd

All rights reserved

RELATIVITY THEORY
OF
PROTONS AND ELECTRONS

by

SIR ARTHUR EDDINGTON
M.A., D.Sc., LL.D., F.R.S.

Plumian Professor of Astronomy and Experimental Philosophy in the University of Cambridge

CAMBRIDGE
AT THE UNIVERSITY PRESS
1936

PRINTED IN GREAT BRITAIN

PREFACE

In this book I have endeavoured to give a connected account of a series of investigations in the borderland between relativity theory and quantum theory. It begins where my earlier book, *The Mathematical Theory of Relativity*, leaves off—at the point where in our survey of nature we encounter the phenomenon of atomicity. To our gross senses matter seems continuous, and it has been treated as continuous in the usual theory of relativity. Experiment has, however, taught us that it is composed of multitudes of units, and the theory is here extended to throw light on the existence and properties of these units.

The central problem is to ascertain the conditions which fix the amount of mass and electric charge carried by protons and electrons. The present researches will probably be associated in the minds of many readers with the number 137; this is one of four numerical constants of nature for which the theory predicts definite values. Another fundamental constant, found to be $2.136.2^{256}$ (approximately $3{\cdot}150.10^{79}$), can be described as the number of protons and electrons in the universe; but its practical importance is that its square root enters into the ratio of the electrical to the gravitational force between a proton and electron. The result of these determinations is that there are no arbitrary constants left in the scale of relations of natural phenomena.

Besides giving concrete results of this kind, the theory has, I hope, thrown light on some of the obscure points in quantum theory and helped to deepen its foundations. I have sought a harmonisation, rather than a unification, of relativity and quantum theory. I do not set out to obtain an all-embracing formula; but the investigation shows in detail how to combine the conceptions of the two theories in the solution of specific problems, which would be outside the range of either theory separately.

The theory, as it was being developed, was published from time to time in the *Proceedings of the Royal Society* (**121**, p. 524; **122**, p. 358; **126**, p. 696; **133**, pp. 311, 606; **134**, p. 524; **138**, p. 17; **143**, p. 327; **152**, p. 253) and the *Journal of the London Mathematical Society* (**7**, p. 58; **8**, p. 142) between 1928 and 1935. But it has become increasingly difficult to deal with it in fragments. Much of it (including practically the whole of Chapters VI, XI and XVI) is now published for the first time. The new results of a practical kind include the theory of the Stern-Gerlach experiment, the theory of Bond's correction $\frac{136}{137}$ to e/m, and the direct calculation of the number of particles in the universe.

In so extensive a work I cannot expect that serious mistakes have been entirely avoided. But now that the theory can be viewed as a whole, I think the reader will be convinced that there is a practicable way of progress along the lines I have attempted. I hope therefore that he will see in the imperfections of this book an opportunity for developing, not an excuse for dismissing, the subject which it sets forth.

Prof. G. Lemaître and Prof. G. F. J. Temple have kindly read the book in proof. Their interest and criticism has encouraged me in the development of the theory, and I now owe them a further debt for many helpful suggestions.

A. S. E.

CAMBRIDGE,
June 1936.

CONTENTS

INTRODUCTION *page* 1

PART I. WAVE-TENSOR CALCULUS

CHAPTER I Tensors and Matrices 13

II The Sixteenfold Frame 20

III The Resolution of Matrices 34

IV Space Vectors 50

V The Simple Wave Equation 62

VI Reality Conditions 75

VII Strain Vectors and Phase Space 92

VIII The Differential Wave Equation 115

IX The Hydrogen Atom 140

X Double Wave Vectors 154

PART II. PHYSICAL APPLICATIONS

XI The Riemann-Christoffel Tensor 179

XII The Mass-ratio of the Proton and Electron 212

XIII Standing Waves 229

XIV The Cosmical Problem 256

XV Electric Charge 281

XVI The Exclusion Principle 308

INDEX 331

INTRODUCTION

0·1. In 1928, P. A. M. Dirac made a bridge between quantum theory and relativity theory by his linear wave equation of the electron.† This is the starting point of the development of relativity theory treated in this book.

Previously there had been three principal stages of progress, namely Einstein's special theory (1905), his general theory (1915), and Weyl's theory of relativity of gauge (1918). Summarising the state of the theory in 1923, I wrote‡

> We offer no explanation of the occurrence of electrons or of quanta; but in other respects the theory appears to cover fairly adequately the phenomena of physics. The excluded domain forms a large part of modern physics, but it is one in which all explanation has apparently been baffled hitherto. The domain here surveyed covers a system of natural laws fairly complete in itself and detachable from the excluded phenomena, although at one point difficulties arise since it comes into close contact with the problem of the nature of the electron.

Relativity theory was in fact as comprehensive and as logically complete as a purely macroscopic theory had any right to be. The next important step must be an extension to cover microscopic phenomena, or a unification with existing microscopic theories.

Microscopic physics was the province of quantum theory; but in 1923 this was little more than a collection of empirical rules which led to no coherent outlook. The "new quantum theory" began with Heisenberg's researches in 1925, and with the aid of many contributors it reached soon afterwards the current form generally called wave mechanics. The conditions were becoming ripe for a unification with macroscopic relativity theory.

To say that Dirac's wave equation was the first connecting link gives only a partial idea of its importance. *It was a challenge to those who specialised in relativity theory.* Dirac's object was to obtain a form of equation (fulfilling certain requirements of quantum theory) which should be invariant for rotations and Lorentz transformations. We had claimed to have in the tensor calculus an ideal tool for dealing with all forms of invariance and covariance. But instead of using the orthodox tool Dirac proceeded by a way of his own, and produced an expression of very unsymmetrical appearance, which he showed to be invariant for the transformations of special relativity theory. Why had this type of invariance eluded the ordinary tensor calculus? As C. G. Darwin put it, "it is rather disconcerting to find

† *Proc. Roy. Soc.* A, 117, 610 (Feb. 1928).
‡ *Mathematical Theory of Relativity*, p. 237.

that apparently something has slipped through the net".† It was Darwin's insistence on this point in private conversation which led me to take up these investigations.

The failure of ordinary tensor calculus to include Dirac's type of invariance is due to the introduction, at an early stage, of a convention whose arbitrariness had already been noticed.‡ The analytical theory of tensors had been applied to physics by identifying its basic vector with a geometrical displacement $(dx)^\mu$. By a change of application, namely by identifying the basic vector with Dirac's four-valued quantity ψ, we obtain a new tensor calculus, here called *wave-tensor calculus*, in which the invariance of the wave equation falls into order. Formulae have to be found for expressing the old tensors (space tensors) in terms of the wave tensors; and this leads to a chain of new developments which have no counterpart in the tensor calculus of ordinary relativity theory.

I was soon convinced that this was the extension of relativity theory for which we had been waiting, and that Dirac's equation was only the beginning of a more far-reaching application of the methods and conceptions of relativity theory to microscopic phenomena. After seven years' work I find the possibilities latent in the new departure still far from exhausted.

Naturally others besides myself were attracted to the new opening. Allowing for divergences in the point of view, my first paper,§ dealing with formal developments including the rudiments of wave-tensor calculus, was perhaps not materially different from several other investigations published about the same time.‖ But a month or two later I came across a clue to the origin of the charge of electrons and protons.¶ The trail has led all round the universe, so that the subject with which I began comes almost at the end of this book (Chapter XV). Ultimately the problem of the origin of charge was found to be inseparable from the problem of the origin of mass. I was thus led into a special field of investigation which, I think, has not been explored by other writers.

Dirac's wave equation has led to important advances in quantum theory; but here we shall be working mainly on the relativity side of the bridge. It is, of course, impossible to treat protons and electrons without introducing a considerable amount of quantum theory. But its subordinate position will be apparent from the fact that the problems treated in Part II of this book are not touched upon in books on quantum theory; they depend essentially on developing the consequences of the relativistic conception.

† *Proc. Roy. Soc.* A, **118**, 654. This paper was of great assistance in my early work.

‡ *Mathematical Theory of Relativity*, p. 49.

§ *Proc. Roy. Soc.* A, **121**, 524.

‖ I think that the most far-reaching, as well as the earliest, paper of this type was by J. v. Neumann, *Zeits. für Physik*, **48**, 868. But I have been more influenced by H. Tetrode, *ibid.* **50**, 346, whose point of view was less unfamiliar to me.

¶ "The Charge of an Electron", *Proc. Roy. Soc.* A, **122**, 358 (Dec. 1928).

0·2. As the work proceeded, it became focused on one problem, namely the origin of the four numerical "constants of nature". Seven fundamental constants are commonly recognised:

m_e the mass of an electron,
m_p the mass of a proton,
e the charge of an electron,
h Planck's constant,
c the velocity of light,
κ the constant of gravitation,
λ the cosmical constant.

Between these we must eliminate our arbitrary units of length, time and mass; we are then left with four purely numerical ratios. The most familiar are the mass-ratio m_p/m_e, and the fine-structure constant $hc/2\pi e^2$; these are found in Chapters XII and XV. The value of κ, i.e. its ratio to a constant of similar dimensions furnished by the other constants, is obtained in Chapter XIV. Also, with the help of the other constants we replace λ by a number N, the "number of particles in the universe", whose theoretical value is found in Chapter XVI. Thus all four constants are obtained by purely theoretical calculation.†

The number of dimensions of space-time might be regarded as a fifth natural constant. Even this number is found to be determined unambiguously by the epistemological principle that we can only observe relations between two entities (§ 16·8). At a much earlier stage (Chapter VI) we prove that a four-dimensional neutral domain necessarily has the signature $3+1$.

So far as I can make out, the values of the constants given by this theory are in full agreement with observation. For three of the four constants the observations are accurate enough to provide a very stringent test. It would have been disconcerting if it had turned out otherwise; but the theory does not rest on these observational tests. It is even more purely epistemological than macroscopic relativity theory; and I think it contains no physical hypotheses—certainly no new hypotheses—to be tested. All that we require from observation is evidence of identification—that the entities denoted by certain symbols in the mathematics are those which the experimental physicist recognises under the names "proton" and "electron". Being satisfied on this point, it should be possible to judge whether the mathematical treatment and solutions are correct, without turning up the answer in the book of nature. My task is to show that our theoretical resources are

† A general account of the principles on which the calculations are based is given in *New Pathways in Science*, Chapter XI.

sufficient and our methods powerful enough to calculate the constants exactly—so that the observational test will be the same kind of perfunctory verification that we apply sometimes to theorems in geometry.

The replacement of four empirical natural constants by calculated numbers implies a unification of theory. In Maxwell's unification of electromagnetism and optics the ratio of the electromagnetic to the electrostatic unit of charge was found to be equal to the velocity of light; similarly in the unification of macroscopic and microscopic theory the macroscopic constants κ, λ are found to be expressible in terms of m_p, m_e, e. The elimination of superfluous constants is an outward sign of the unification achieved; and for this reason I have regarded it as the first goal. But the theory that has taken shape in these investigations should supply a foundation for the treatment of other microscopic problems for which current quantum theory is insufficient. I have not pursued these further developments, partly because they often require a knowledge of the more technical side of quantum theory which I do not possess, and partly because the completion of the calculation of the four natural constants has seemed an appropriate stage at which to assemble the theory into a connected form. By way of exception, I have applied the theory to the Stern-Gerlach experiment (§ 12·8); the result agrees with observation.

0·3. The marriage of relativity theory and quantum theory should be a fruitful union as well as a formal union. In regard to the numerous formal unified theories that have been suggested, we may recall—

> There are nine and sixty ways of constructing tribal lays,
> And every single one of them is right!

But I think they have been inspired by a fundamentally different conception of the problem of unification from that which I shall follow. There is a considerable amount of formal theory in this book; but it has been developed concurrently with the physical theory in Part II; and its progress has been guided as much by the definite applications, in which it was to be used, as by formal considerations.

A *unified theory* does not necessarily mean a *unified formula*. The latter kind of unification is exemplified by the theory of the "Generalised Astronomical Instrument" which combines in a single equation the theory of the altazimuth, meridian circle, prime vertical instrument, equatorial and almucantar.† Such compression appeals more to the mathematician than to the physicist. We do not aim at producing a formula which shall include simultaneously the irregular gravitational fields of general relativity and the quantised energy of an atom. We seek instead the common meeting point from which the specialised developments and approximations appro-

† *Monthly Notices, R.A.S.* **68**, 171.

priate to the gravitational and the atomic problem branch off. The source of quantum phenomena is a degeneracy, or exceptional integrability, which is associated with uniformity and symmetry, and invalidates the assumptions underlying the ordinary theory of macroscopic averages. It is therefore in uniform conditions (spherical space) that the linkage of quantum theory to macroscopic relativity theory must primarily be studied.

For this reason we have not much to do with the formulae of general relativity, though we have much to do with its principles. We generally treat space-time of uniform curvature—either the de Sitter form hyperbolic in the time dimension, or the Einstein form cylindrical in the time dimension, the space being spherical in either case. Thus, on the relativity side, we halt at a stage intermediate between the general theory and the special relativity theory of flat space.

For the same reason, when gauge transformations are employed, the formulae used are those of Weyl's theory, not the author's generalisation of it.† The generalisation would be required if we dealt with space-time of irregular curvature; it coalesces with Weyl's theory in the uniform conditions here considered.

With *flat* space-time we have nothing to do. The theory of space-time will here be developed *pari passu* with the theory of the material systems which occupy it. In this mode of approach the conception of infinite flat space never arises; it could not be brought into the theory except as a limit that might be approached but never attained. But in the same way the conception of definitely empty space never arises; it could not be brought into the theory except as a limit that might be approached but never attained. (By *definitely empty* we mean that the probability of containing a particle or photon is zero.) Our rejection of flat space-time does not depend on the view that definitely empty space has a natural curvature determined by the cosmical constant.‡ Space appears in our theory as the domain of the probability distribution of a particle, so that it is an essential characteristic of space that it is occupied or has a finite *a priori* probability of being occupied; and it is non-controversial that it will have a curvature (or an expectation-value of the curvature) corresponding to the energy tensor (or expectation-value of the energy tensor) of its contents.

0·4. I think it will be found that the theory is purely deductive, being based on epistemological principles and not on physical hypotheses. But it could not be presented in purely deductive form—which would mean, I suppose, that it was treated as an investigation in pure mathematics with a physical *dénouement* in the last chapter. It has seemed essential to

† *Mathematical Theory of Relativity*, Chapter VII, Pt. II.

‡ We are led to reject this view (§ 11·7).

keep the physical applications in mind throughout; for this purpose results must sometimes be anticipated which are not reached by deduction until much later in the book, and interpretations must be employed which are not definitely established until the whole theory is connected together.

This method gives rise to certain difficulties. For mathematical reasons we have to begin with the simplest equations; but these correspond to highly idealised systems to which the ordinary physical conceptions only partially apply. It is indeed obvious that a system must attain a considerable degree of complexity before anything remotely resembling the ordinary method of observation is applicable to it. Consequently the physical ideas can only take shape gradually as we proceed. Space-time of a kind first appears in Chapter IV; but it has the wrong signature, and its scale is much too small. These defects become rectified as the developments in later chapters take us closer to actuality. So with most of the physical conceptions; we have to introduce preliminary notions before the theory is sufficiently advanced for a full treatment. The reader will probably find that many of the difficulties that occur to him in reading the earlier chapters arise from the unnatural conditions postulated in the most elementary equations, and that they resolve themselves automatically when the theory reaches more realistic problems.

Those who are expert in quantum theory should bear in mind that we are proceeding from another starting point from that usually adopted. It would be foreign to my intention of developing the theory as a pure deduction from relativistic principles to transfer conclusions, however widely accepted, from the usual quantum theory which contains a large empirical element. Nevertheless, I make frequent appeals to current quantum theory for three purposes. Firstly, because it contains the *definitions* of the quantities with which I am concerned. It would be impossible to make a theoretical determination of the constant known as the "mass of an electron" without an examination of the equations by which the quantum physicist has chosen to define it. Secondly, where the present theory coalesces with current theory, it is unnecessary to repeat purely analytical investigations which equally apply to either theory. Thirdly, certain results (especially the Exclusion Principle), which cannot be treated until late in the book, have been borrowed from current theory in anticipation.

It may be well to make it clear that although the present theory owes much to Dirac's theory of the electron, to the general coordination of quantum theory achieved in his book *Quantum Mechanics*, and to the many contributions of himself and others on these lines, it is not "Dirac's Theory"; and indeed it differs fundamentally on most points which concern relativity. It is definitely opposed to what has commonly been called "relativistic

quantum theory", which, I think, is largely based on a false conception of the principles of relativity theory.

Atomic nuclei and free neutrons are outside the scope of this book. I see no reason to fear that they will not fall into place in the theory; but I have not developed any ideas on this point far enough to be worth recording. In the main the theory of radiation has also been excluded; but there are three short references (§§ 9·5, 14·3, 16·3) which show how it might be approached in the present treatment.

0·5. The division of the book into two parts, the one ostensibly treating the auxiliary mathematical calculus and the other the physical applications, is only a rough separation. Physical interpretations are considered as early as possible; but with the introduction of double wave tensors in Chapter x the relation of the mathematics to the physics changes considerably. Instead of starting with the mathematical result and interpreting it as far as possible physically, we start with the physical problem and formulate it mathematically. The auxiliary mathematical development still continues; but it is now guided by the character of the physical problems for which its aid is required.

The student of relativity theory may well feel a grievance at the turn which the auxiliary mathematics has taken. The macroscopic theory seemed to indicate that Differential Geometry was the key to world-structure. After being at pains to acquire some familiarity with this subject, we find that all the new advances depend upon modern Algebra. The algebra required in the present book is developed practically *ab initio*—by old-fashioned methods which, I fear, betray my limitations as an algebraist, though they may make the theory more accessible to those most interested. Let me freely admit that ability to use the more powerful modern algebraic methods would be an immense advantage in handling these problems. For the kind of algebra chiefly required I have found most helpful C. C. MacDuffee, *The Theory of Matrices*.†

A few remarks on terminology, etc. may be useful. I would direct special attention to the *limited* use of the summation convention (p. 22), to my unorthodox use of the term "algebraic" (p. 20), and to the change in the order of writing the suffixes of the Riemann-Christoffel tensor (p. 181). I would emphasise that wave analysis is a method, not a theory, and may be applied to *any* physical tensor; therefore statements about the physical meaning of the various products of wave analysis necessarily refer to some special application (singled out by custom) and are not of general validity.

The term *universe* is used so often as perhaps to suggest megalomania. It is really the opposite of megalomania, for it takes the place of *infinity* in elementary wave mechanics. Mathematically it is much easier to treat a

† *Ergebnisse der Mathematik und ihrer Grenzgebiete* (Julius Springer, 1933).

whole universe than part of one—the universe being, of course, idealised to accord with the simple conditions postulated in elementary problems. It is more elementary to suppose that the uniform conditions continue indefinitely than to terminate them by a physical barrier; supernatural barriers are often misleading, and should be avoided if possible. In the earlier chapters there is sometimes a difficulty in deciding whether our equations refer to an electron or to the universe. But the fact is that the electrons treated in ordinary elementary quantum theory are very much like the universe—only bigger. They are said to be "infinite plane waves". No doubt it is intended that they shall be replaced by waves of more reasonable dimensions in practical approximations; but this applies also to our theory. For applications in which a millimetre is a good enough approximation to infinity, it is *a fortiori* a good enough approximation to 400 megaparsecs.

By a *particle* I mean, not a classical particle, but a conceptual entity whose probability distribution is specified by a wave function. At different stages in this book, different applications of wave analysis are made; and the corresponding particles have different properties. In the earlier chapters the particles are rudimentary protons and electrons existing in the rudimentary space-time there treated. They gradually develop into recognisable electrons and protons in macroscopic space-time, when the theory is extended far enough to introduce the observable relations by which protons and electrons are known experimentally. The reader should therefore not be surprised to find that initially the positive and negative particles have completely symmetrical properties; that is merely another illustration of the fact that the most elementary equations imply highly idealised conditions to which the ordinary conceptions of physics only partially apply.

0·6. Those who have followed the progressive development of the theory during the last eight years may desire a comparison of the present revised theory with earlier versions. The papers are numbered for reference as follows:†

I. "A Symmetrical Treatment of the Wave Equation", *R.S.* **121**, 524, 1928
II. "The Charge of an Electron", *R.S.* **122**, 358, 1928.
III. "The Interaction of Electric Charges", *R.S.* **126**, 696, 1930.
IV. "The Properties of Wave Tensors", *R.S.* **133**, 311, 1931.
V. "The Value of the Cosmical Constant", *R.S.* **133**, 605, 1931.
VI. "The Mass of a Proton", *R.S.* **134**, 524, 1931.
VII. "Sets of Anticommuting Matrices", *M.S.* **7**, 58, 1931.
VIII. "Theory of Electric Charge", *R.S.* **138**, 17, 1932.
IX. "The Factorisation of E-Numbers", *M.S.* **8**, 142, 1933.
X. "The Masses of the Proton and Electron", *R.S.* **143**, 327, 1933.
XI. "The Pressure of a Degenerate Gas, and Related Problems", *R.S.* **152**, 253, 1935.

† *R.S.* refers to *Proc. Roy. Soc.* and *M.S.* to *Journ. Lond. Math. Soc.*

I will begin with a definite withdrawal. In I it was suggested that the adjoint tetrads E_1, E_2, E_3, E_4 and $E_{15}, E_{25}, E_{35}, E_{45}$ correspond to electrons of opposite spin, and this misjudgment persisted in III. The present view (§§ 4·3, 4·4) of the rôle of E_5 was first reached in IV. I think there is no other point on which I went so completely astray; the other lines of development begun in this series of papers, though sometimes requiring substantial amendment, contain advances which in principle have been retained. The papers fall into three groups:

(*a*) *Auxiliary Mathematics* (I, IV, VII, IX). The E-symbols were at first defined so that $E_\mu^2 = 1$; the present notation, with $E_\mu^2 = -1$, begins in IV. In IV the mathematical development remains satisfactory; but the physical interpretation was confused, because a degenerate wave tensor was used in a context where the later results substitute a non-degenerate wave tensor. The proofs of two important theorems in VII and IX, namely the composition of a pentad by three imaginary and two real matrices, and the standard forms of pure wave tensors, have been found to be imperfect. Amended proofs are given in §§ 3·5, 5·5.

(*b*) *Electric Charge* (II, III, VIII). Paper II has been affected less than most of the early papers by subsequent progress, and can be regarded as substantially correct so far as it goes. Gaunt's form of the matrix coefficient of the Coulomb energy, employed in the paper, is now obsolete; but this scarcely affects the investigation at the stage concerned. The factor 136 was changed to 137 in subsequent papers; but the difference is a question of definition (§ 15·9). Paper III represents an interim stage in a complicated investigation, and has the defects of an interim report. Progress remained unsatisfactory until the interchange energy was associated with the operator P instead of with E_5. The theory in VIII is substantially the same as the second of the two methods given in this book (§ 15·7).

(*c*) *Origin of Mass* (V, VI, X, XI). With regard to V and VI, which were preliminary papers, I need only say that the present theory follows the ideas there suggested. In X the argument now replaced by the formula $R_0^2 = RR'$ was unsatisfactory; otherwise the changes are mainly of the nature of amplification. Paper XI is practically up to date; but a numerical change was made necessary by the discovery of an inconsistency of a factor 2 in current quantum theory (§ 9·6).

The various lines of investigation were very much interlocked; a backward state of one prevented progress in the others. Thus the whole work reached completion as one unit. The cosmical problem treated in XI was the last item on the main programme; and, after it was solved, there was not much difficulty in supplying the remaining investigations needed to fit together all the material. The investigations in the published papers

are, of course, considerably altered in form now that they are connected to a homogeneous theory instead of to fragments of current theory partially modified to suit the growth of ideas.

It was found in III that the theory offers an explanation why one dimension of the world differs from the other three; but, except for this, little attention was paid to reality conditions in the series of papers. This was deliberate; because it seemed premature to try to formulate reality conditions before the main lines of connection of the analytical theory with observational phenomena were settled. It was not until after the last of the published papers that I took up the problem and reached the reality conditions formulated, and used extensively, in this book.

Strain vectors first appeared in X, but were to some extent anticipated in VIII. The more extended use of strain vectors, and the systematic discrimination between internal and external wave functions, is a feature of the new treatment. Other portions of the theory scarcely touched on in the published papers are §§ 8·4–9·1, 10·4–11·9, 12·6–12·8, 15·8–16·9.

At one time I laid stress on a suggestion (due to Zanstra) that the packing ratio in helium is an approximation to $\frac{137}{136}$. It seemed likely that the binding of particles in a *rigid* nucleus might be represented as the loss of one degree of freedom of the double wave functions, with a corresponding reduction of the energy required for statistical equilibrium. Recent atomic weights make the packing ratio less close to $\frac{137}{136}$ than was at one time supposed; but in any case the theory of a rigid nucleus was not expected to apply to helium exactly. I still regard the suggestion as plausible; but as my investigations have not dealt with nuclear structure, the question remains in suspense.

PART I

WAVE-TENSOR CALCULUS

CHAPTER I

TENSORS AND MATRICES

1·1. Linear Transformations.

A physical system may be described in many alternative ways. Different systems of coordinates may be used for specifying its position; different systems of units may be used for the measurement of mass, length, time; and so on. Accordingly our attention is directed to the problem of comparing systems of description in which there is a one-to-one correspondence between quantities A, B, C, ... occurring in one description and quantities A', B', C', ... occurring in another description.

The description commonly includes sets of associated quantities which are regarded as "components" of a single entity, e.g. the three components of a force. We then have a correspondence between an array of n quantities A_μ in one description and A_μ' in another description ($\mu = 1, 2, \ldots n$).

We proceed at once to a special case of great importance, viz. when A_μ' is given by a linear transformation of A_μ

$$\left.\begin{aligned} A_1' &= q_{11}A_1 + q_{12}A_2 + \ldots + q_{1n}A_n \\ A_2' &= q_{21}A_1 + q_{22}A_2 + \ldots + q_{2n}A_n \end{aligned}\right\} \qquad (1\cdot11)$$

etc. Using the summation convention of the tensor calculus, these formulae are written more compactly

$$A_\sigma' = q_{\sigma\mu}A_\mu \qquad (1\cdot12)$$

and the transformation, or change of description, is described as $A_\sigma \to q_{\sigma\mu}A_\mu$.

The array of coefficients $q_{\sigma\mu}$ defines the change of the system of description, so far as the characteristic A_μ is concerned. Linear transformations possess the Group property; that is to say, the resultant of a succession of linear transformations is a linear transformation. Thus we can have a set of systems of description such that, in passing from any one description to any other, the transformation of A_μ is always linear. When for all systems of description contemplated the transformation of A_μ is linear, A_μ is called a *tensor*.

By solving equations (1·11) we can find $A_1, A_2, \ldots$ in terms of $A_1', A_2', \ldots$. The resulting formulae are linear and may be written

$$A_\sigma = q_{\sigma\mu}'A_\mu'. \qquad (1\cdot13)$$

The array of coefficients $q_{\sigma\mu}'$ defines the *inverse* transformation to that defined by $q_{\sigma\mu}$.

If B_μ is another array of n quantities occurring in the description of the physical system, and in the change of description in which $A_\sigma \to q_{\sigma\mu}A_\mu$

$$B_\sigma \to q_{\sigma\mu}B_\mu, \qquad (1\cdot14)$$

B_μ is said to be a *tensor of the same kind* as A_μ (or to be *cogredient* with A_μ).

If C_μ is another array of n quantities occurring in the description, and in the change of description in which $A_\sigma \to q_{\sigma\mu} A_\mu$

$$C_\sigma \to q_{\mu\sigma}{}' C_\mu, \tag{1·15}$$

C_μ is said to be a *tensor of opposite kind* to A_μ (or to be *contragredient* to A_μ). Note the inversion of the order of the suffixes of q'.

From A_μ and C_μ we may form an array of n^2 quantities $A_\mu C_\nu$ which follows the transformation law

$$A_\sigma{}' C_\tau{}' = q_{\sigma\mu} A_\mu q_{\nu\tau}{}' C_\nu = q_{\sigma\mu} q_{\nu\tau}{}' A_\mu C_\nu. \tag{1·16}$$

If $T_{\mu\nu}$ is an array of n^2 quantities occurring in the description, and in the change of description in which $A_\mu \to q_{\sigma\mu} A_\mu$

$$T_{\sigma\tau} \to q_{\sigma\mu} q_{\nu\tau}{}' T_{\mu\nu} \tag{1·17}$$

(i.e. if it is transformed in the same way as $A_\mu C_\nu$), then $T_{\mu\nu}$ is said to be a *mixed tensor of the second rank* of the class A_μ.

Tensor properties do not necessarily depend on the physical nature of the entity that is being described; they depend on the variety of descriptions which we admit. For example, the statement that B_μ is a tensor of the same kind as A_μ announces a limitation of the variety of description contemplated; for there can be no *compulsion* to change our description of one physical feature of the system when the description of another feature is changed. But unless there is some systematic plan underlying our descriptions it will be impossible to assert any general laws governing the quantities occurring in the descriptions.

For example, the strength of the wind is sometimes described by a number of dynes per square centimetre and sometimes by a number on the Beaufort scale. We cannot expect to find exact equations (relating our measures of the strength of the wind to other meteorological characteristics) applicable to *both* codes of measurement. By taking the wind strength to be a tensor of the class of tensors used for describing other meteorological characteristics we rule out one or other description—not as illegitimate, but as unsuited to the purpose we have in mind, viz. to express the regularities underlying natural phenomena by mathematical equations governing the quantities which occur in our descriptions of the phenomena.

1·2. Space Tensors and Wave Tensors.

When the change of system of description includes a change of coordinates from (x_1, x_2, x_3, x_4) to (x_1', x_2', x_3', x_4'), an infinitesimal coordinate difference dx_μ is transformed according to the formula

$$dx_1' = \frac{\partial x_1'}{\partial x_1} dx_1 + \frac{\partial x_1'}{\partial x_2} dx_2 + \frac{\partial x_1'}{\partial x_3} dx_3 + \frac{\partial x_1'}{\partial x_4} dx_4 \tag{1·21}$$

etc. This may be written in the form (1·15)

$$dx_\sigma' = q_{\mu\sigma}' dx_\mu \quad (q_{\mu\sigma}' = \partial x_\sigma'/\partial x_\mu). \tag{1·22}$$

Thus every change of description contemplated as admissible corresponds to a linear transformation of dx_μ. Accordingly dx_μ is a tensor; we call it a *displacement vector*.

This is the basic tensor of the class of tensors used in the ordinary tensor calculus. Displacement vectors and all tensors of the same kind are called contravariant vectors; tensors of opposite kind are called covariant vectors. Mixed tensors of the same class are defined as in (1·17); and more generally tensors of higher rank with 16, 64, 256, ... components are introduced, their transformation laws being

$$A'^{\alpha\beta\cdots}_{\gamma\delta\cdots} = q_{\mu\alpha}' q_{\nu\beta}' \cdots q_{\gamma\sigma} q_{\delta\tau} \cdots A^{\mu\nu\cdots}_{\sigma\tau\cdots}. \tag{1·23}$$

We shall call this class of tensors *space tensors*.

Thus, although the theory of tensors belongs primarily to the algebraic theory of transformations, it has usually been linked to geometry by identifying the basic tensor of the algebraic scheme with a geometrical displacement or coordinate difference dx_μ. *We shall here discard this special linkage.* We shall introduce another class of tensors called *wave tensors*, derived from a basic contravariant wave vector χ_μ in the same way that the space tensors are derived from the basic contravariant space vector dx_μ.

For the moment we leave the basic wave vector unidentified. But at a certain point in the development of the system of wave tensors, we shall be able to side-step into a new class of tensors. On examining the properties of the new tensors we shall find that they can be identified with space tensors. Thus the wave-tensor calculus leads up to the ordinary space-tensor calculus and includes it as a side branch; but its greater comprehensiveness fits it to deal with certain entities in modern quantum theory which are not describable by space tensors.

The basic wave vector will be identified in Chapter v. It turns out to be the four-valued wave symbol introduced into physics by P. A. M. Dirac in his linear wave equation of the electron. Vectors of this class cannot be reached from the ordinary calculus of space tensors, which does not begin far enough back. Our plan accordingly is to begin with these vectors, and lead up to the ordinary space vectors at a later stage.

1·3. Chain Multiplication.

Let $A_\mu{}^\nu$, $B_\mu{}^\nu$ be two mixed tensors of the second rank. Having regard to the summation convention we recognise four different products

$$A_\mu{}^\nu B_\sigma{}^\tau, \quad A_\mu{}^\nu B_\nu{}^\tau, \quad A_\mu{}^\nu B_\sigma{}^\mu, \quad A_\mu{}^\nu B_\nu{}^\mu. \tag{1·31}$$

The first is the *outer product*, and the fourth is the *inner* or *scalar product*.

The second and third are called *matrix products* and are denoted by AB and BA respectively.

Matrix products are formed by chain multiplication, i.e. the second suffix of one factor is repeated as the first suffix of the succeeding factor (the repetition introducing a summation in accordance with the summation convention). The product $A_\mu{}^\nu B_\nu{}^\tau$ is of this form. $A_\mu{}^\nu B_\sigma{}^\mu$ is not a chain product as it stands; but it becomes one if it is rewritten as $B_\sigma{}^\mu A_\mu{}^\nu$.

On the understanding that chain multiplication is the only kind of multiplication admitted, no suffixes need appear in the formulae, since the reader can always supply appropriate suffixes when required. Thus the product of a number of double-suffixed quantities is written

$$P = ABCD, \tag{1·321}$$

which stands for

$$P_\mu{}^\nu = A_\mu{}^\alpha B_\alpha{}^\beta C_\beta{}^\gamma D_\gamma{}^\nu. \tag{1·322}$$

This rule of multiplication is the distinctive feature of the matrix calculus. The notation is so useful that we cannot afford to do without it. Nevertheless matrix calculus suffers from being more limited than tensor calculus; and we often want to introduce outer and scalar products and other combinations for which matrix calculus provides no notation. This necessitates resorting to various awkward shifts, and occasionally reverting to the full suffixed expressions.

Chain multiplication does not contemplate quantities with more than two suffixes. We shall at first limit the term "matrix" to two-suffixed quantities representing two-dimensional arrays. Technically one-dimensional arrays are also matrices, but it would probably be confusing to include them. One-dimensional arrays will here be called vectors, even when no question of transformation properties arises. The term implies very little restriction so long as we do not specify the *kind* of vector.

Chain multiplication cannot be carried beyond a vector, so that vectors can only occur at the beginning or end of a matrix product. We shall distinguish initial vectors by an asterisk, final vectors being unmarked. This notation allows us to reintroduce outer multiplication to a limited extent. The rule is that, if it is impossible to interpret two symbols in juxtaposition as a chain product, they are to be interpreted as an outer product. Thus if ψ_μ is a vector, the expressions

$$A\psi B, \quad A\psi^* B$$

are interpreted as

$$(A\psi) \times B, \quad A \times (\psi^* B),$$

where the symbol $\times$ indicates outer multiplication, chain multiplication being impossible after a final vector or before an initial vector. Or, with suffixes,

$$A\psi B = A_\mu{}^\nu \psi_\nu B_\sigma{}^\tau, \quad A\psi^* B = A_\mu{}^\nu \psi^\sigma B_\sigma{}^\tau.$$

In particular, we have the following notation which is of great importance

$$\left.\begin{aligned} &\psi\chi^* \text{ denotes the outer product } \psi_\mu \chi^\nu, \\ &\chi^*\psi \text{ denotes the scalar product } \chi^\mu \psi_\mu . \end{aligned}\right\} \tag{1·33}$$

The asterisk is a substitute for suffix indications, and is dropped when the suffixes are inserted.

A feature of matrix multiplication is that it is non-commutative; that is to say

$$BA \neq AB. \tag{1·34}$$

It is to be remembered that the non-commutation only arises through the omission of suffixes; when suffixes are inserted in BA, the factors commute as usual. Thus

$$B_\mu{}^\alpha A_\alpha{}^\nu = A_\alpha{}^\nu B_\mu{}^\alpha. \tag{1·35}$$

Since the suffixes are often omitted, we can no longer depend on discriminating contravariant from covariant vectors by the upper and lower positions of the suffixes. There would be little advantage in retaining a method of discrimination which only worked spasmodically. Accordingly, we shall in future generally write all wave tensor suffixes in the lower position.

1·4. Transformation Laws of Wave Tensors.

In § 1·1 we introduced three kinds of tensors of the class A_μ with transformation laws (1·14), (1·15) and (1·17) respectively. The formulae may be written as

$$B_\sigma{}' = q_{\sigma\mu} B_\mu, \quad C_\sigma{}' = C_\mu q_{\mu\sigma}{}', \quad T_{\sigma\tau}{}' = q_{\sigma\mu} T_{\mu\nu} q_{\nu\tau}{}'.$$

The products, as here written, are all chain products, so that the suffixes may be omitted and we have

$$B' = qB, \quad C^{*\prime} = C^* q', \quad T' = qTq'. \tag{1·41}$$

Further by (1·12) and (1·13)

$$A_\sigma{}' = q_{\sigma\mu} A_\mu, \quad A_\mu = q_{\mu\tau}{}' A_\tau{}'.$$

Therefore

$$A_\sigma{}' = q_{\sigma\mu} q_{\mu\tau}{}' A_\tau{}'. \tag{1·421}$$

But

$$A_\sigma{}' = \delta_{\sigma\tau} A_\tau{}', \tag{1·422}$$

where $\delta_{\sigma\tau}$ is the substitution operator, viz.

$$\left.\begin{aligned} \delta_{\sigma\tau} &= 1, \quad \text{if } \sigma = \tau \\ &= 0, \quad \text{if } \sigma \neq \tau \end{aligned}\right\}. \tag{1·43}$$

Since $A_\tau{}'$ is an arbitrary array of four numbers, it follows from (1·421) and (1·422) that

$$q_{\sigma\mu} q_{\mu\tau}{}' = \delta_{\sigma\tau}. \tag{1·44}$$

The left-hand side is a chain product; we can therefore drop the suffixes, obtaining

$$qq' = \delta. \tag{1·45}$$

In matrix calculus δ has the algebraic properties of the number 1. For, if S is any matrix

$$\delta_{\mu\alpha}S_{\alpha\nu}=S_{\mu\nu}, \quad S_{\mu\alpha}\delta_{\alpha\nu}=S_{\mu\nu},$$

so that, dropping suffixes, $\delta S=S$, $S\delta=S$.

Accordingly δ is called the *unit matrix*; and since it is equivalent to the number 1 in matrix calculus, we shall often denote it by 1, or with suffixes $(1)_{\mu\nu}$. Then (1·45) becomes $qq'=1$. Thus q' may be called the reciprocal of q, and it will sometimes be written as q^{-1}.

The formulae (1·41) and (1·45) constitute the principal transformation formulae in wave-tensor calculus. Summarising our results for reference, and changing to the notation which we shall usually employ, we have the following classification and nomenclature:

Covariant (final) wave vectors

$$\psi'=q\psi. \tag{1·461}$$

Contravariant (initial) wave vectors

$$\chi^{*\prime}=\chi^{*}q'. \tag{1·462}$$

Mixed wave tensors

$$T'=qTq', \tag{1·463}$$

with

$$qq'=1. \tag{1·464}$$

These would reduce to the transformation laws of the ordinary tensor calculus if we set

$$q_{\sigma\mu}=\partial x_\mu/\partial x_\sigma', \quad q_{\sigma\mu}'=\partial x_\mu'/\partial x_\sigma. \tag{1·47}$$

But, as already explained, the wave tensors are not linked to geometry in this way, and (1·47) does not apply. For a transformation of wave tensors, any matrix which has a reciprocal may be used as q; that is to say, the corresponding transformation will give a new description which is included in the whole group of descriptions contemplated.

A matrix which has no reciprocal is said to be *singular*. A singular matrix may be regarded as a generalisation of the algebraic number 0 in much the same way that the unit matrix is a generalisation of the number 1; but there are infinitely many different singular matrices. As q approaches a singular value, one or more elements of its reciprocal q' tend to infinity; singular matrices q are therefore excluded in the foregoing transformation theory.

1·5. Initial and Final Wave Vectors.

The terms "initial" and "final", applied to wave vectors, define their behaviour in regard to chain multiplication, and do not necessarily describe their actual position in the sequence of factors (cf. (1·33)). As far as Chapter VI (inclusive) the initial vectors will be contravariant and the final vectors covariant. But it must not be supposed that this is a general rule, or that the asterisk is a symbol for contravariance.

In order to express the covariant transformation law (1·461) in a form appropriate for an initial covariant vector, we introduce a matrix $\bar{q}$ which is the *transpose* of q, obtained by interchanging rows and columns; thus

$$\bar{q}_{\alpha\beta} = q_{\beta\alpha}. \tag{1·51}$$

Then (1·461) stands for

$$\psi_\alpha{}' = q_{\alpha\beta}\psi_\beta = \bar{q}_{\beta\alpha}\psi_\beta = \psi_\beta \bar{q}_{\beta\alpha}.$$

Hence, dropping suffixes, $\psi^{*\prime} = \psi^* \bar{q}$. Treating (1·462) similarly, we have the transformation laws:

Initial covariant wave vectors

$$\psi^{*\prime} = \psi^* \bar{q}, \tag{1·521}$$

Final contravariant wave vectors

$$\chi' = \bar{q}'\chi. \tag{1·522}$$

The outer product $\psi\phi^*$ of two covariant wave vectors ψ, ϕ is a covariant wave tensor S. Using (1·461) and (1·521) we obtain the transformation law:

Covariant wave tensors $$S' = qS\bar{q}. \tag{1·53}$$

These formulae will not be required until Chapter VII.

CHAPTER II

THE SIXTEENFOLD FRAME

2·1. Symbolic Calculus.

For our physical applications the significance of a matrix is embodied, not so much in its representation as an array of numbers, as in its non-commutative multiplication property (1·34). Most, if not all, of the properties of matrices which make them suitable for describing the conditions and activities of the physical universe are also possessed by general symbols endowed with the same non-commutative properties.

We shall therefore develop a calculus containing a number of symbols which do not obey the commutative law of multiplication, but obey the other elementary laws of algebra. The following definitions are adopted:

A symbol which commutes with every symbol in the calculus will be called an *algebraic number*.

The number 1 is defined to be a symbol which satisfies

$$1.E = E.1 = E,$$

where E is any symbol in the calculus. From the definition of 1 the definitions of other algebraic numbers follow in the usual way. In particular i is defined to be a symbol satisfying

$$iE = Ei, \quad iiE = -E.$$

The underlying idea is that a symbol has no properties except such as are manifested by it in the operations of the calculus in connection with which it is used. Its nature lies in its behaviour; it has no intrinsic nature. Therefore if a symbol behaves like the number 1 in every possible operation of the calculus, it *is* the number 1. If our calculus is afterwards extended by the introduction of additional symbols or operations which give a further opportunity for discriminating behaviour, some of the symbols originally counted as algebraic may cease to be algebraic. We may regard "algebraic" as a *relative* characteristic depending on the range of symbols which constitutes our calculus.

I have here deviated from the terminology in pure mathematics, where it is customary to give a much wider meaning to the term "algebraic". But I think that most readers of a physical treatise will naturally understand "algebra" to mean "ordinary algebra"; and therefore the distinction between quantities which obey the rules of ordinary algebra (including the commutative law of multiplication) and those which do not is most intelligibly described by the adjectives "algebraic" and "non-algebraic".

2·2. Complete Orthogonal Sets.

Let E_1, E_2, E_3, E_4 be four symbols which satisfy

$$E_\mu{}^2 = -1, \quad E_\mu E_\nu = -E_\nu E_\mu \quad (\mu, \nu = 1, 2, 3, 4;\ \mu \neq \nu). \tag{2·21}$$

That is to say, the symbols are four mutually anticommuting square roots of -1. We shall find in § 3·2 that there exist matrices which satisfy (2·21), so that we need have no qualms as to the legitimacy of postulating such symbols.

When we are given an even number of anticommuting square roots of -1, we can always find an additional anticommuting square root, making the total number odd. Let

$$iE_5 = E_1 E_2 E_3 E_4. \tag{2·22}$$

We have

$$\begin{aligned}(iE_5)^2 &= E_1 E_2 E_3 E_4 E_1 E_2 E_3 E_4 \\ &= E_1 E_1 E_2 E_2 E_3 E_3 E_4 E_4,\end{aligned}$$

since the rearrangement of order involves six jumps of a symbol over a different symbol, and each jump reverses the sign of the expression by (2·21). Hence

$$(iE_5)^2 = E_1{}^2 E_2{}^2 E_3{}^2 E_4{}^2 = (-1)(-1)(-1)(-1) = 1,$$

so that $E_5{}^2 = -1$. We can verify similarly that $E_1 E_5 = -E_5 E_1$, etc.

Thus we have five symbols satisfying (2·21). Both equations of (2·21) are included in the form

$$\tfrac{1}{2}(E_\mu E_\nu + E_\nu E_\mu) = -\delta_{\mu\nu} \quad (\mu, \nu = 1, 2, 3, 4, 5), \tag{2·23}$$

where $\delta_{\mu\nu}$ is the symbol defined in (1·43).

Any product formed by repeated multiplication of E_1, E_2, E_3, E_4 can be reduced to the form $\pm E_1{}^p E_2{}^q E_3{}^r E_4{}^s$, since in collecting the factors the alteration of order can at most change the sign of the product. Also, since $E_\mu{}^2 = -1$, $E_\mu{}^\rho$ reduces to $\pm E_\mu$ or ± 1. Thus, disregarding sign, the product reduces to one or other of sixteen forms:

$$1,\ E_\mu,\ E_\mu E_\nu,\ E_\mu E_\nu E_\sigma,\ E_1 E_2 E_3 E_4 \quad (\mu, \nu, \sigma = 1, 2, 3, 4;\ \mu \neq \nu \neq \sigma). \tag{2·24}$$

Multiplying (2·22) by E_1, we have

$$iE_1 E_5 = E_1{}^2 E_2 E_3 E_4 = -E_2 E_3 E_4,$$

so that by using E_5 the triple products can be reduced to double products. Disregarding factors ± 1, $\pm i$, the forms (2·24) are equivalent to the sixteen forms

$$i,\ E_\mu,\ E_\mu E_\nu \quad (\mu, \nu = 1, 2, 3, 4, 5;\ \mu \neq \nu). \tag{2·25}$$

As here written they are all square roots of -1; since

$$(E_\mu E_\nu)^2 = E_\mu E_\nu E_\mu E_\nu = -E_\mu E_\mu E_\nu E_\nu = -E_\mu{}^2 E_\nu{}^2 = -1.$$

A linear function of the sixteen expressions (2·25) with algebraic coefficients (real or complex) will be called an *E-number*. We see that the operations of addition, subtraction and multiplication applied to E-numbers will always yield E-numbers. In virtue of this property the sixteen expres-

sions are said to constitute a *complete set*. Similarly in algebra the symbols 1 and i constitute a complete set, since the operations of addition, subtraction and multiplication applied to complex numbers always yield complex numbers.†

For reasons which will appear later the sixteenfold complete set here introduced is called an orthogonal set.

The E-numbers are a particular case ($n=4$) of Clifford's numbers,‡ which are formed analogously from any even number n of independent anticommuting square roots of -1. Since the E-numbers, or their equivalent matrices, play a fundamental part in the physical theory which we shall develop, the theory is dependent on the choice $n=4$ which we make at the outset. This choice will ultimately be justified in § 16·8, where it is shown that it is imposed by the epistemological principles involved in the conception of measurement.

2·3. Notation of the E-symbols.

We shall write $$E_{\mu\nu}=E_\mu E_\nu \quad (\mu,\nu=1,2,3,4,5;\ \mu\neq\nu). \tag{2·31}$$

For uniformity we also give the original five symbols an alternative double-suffix notation, viz. $$E_\mu=E_{0\mu}=-E_{\mu 0}. \tag{2·32}$$

Then the sixteen expressions (2·25) which constitute the complete set become $$i,\ E_{\mu\nu} \quad (\mu,\nu=0,1,2,3,4,5;\ \mu\neq\nu). \tag{2·33}$$

By (2·31) and (2·32) we have in all cases $E_{\mu\nu}=-E_{\nu\mu}$. In making up the complete set of sixteen symbols it is arbitrary whether we employ $E_{\mu\nu}$ or $E_{\nu\mu}$. It would, of course, be redundant to include both.

By using (2·21) and (2·22) we find the following general rules of multiplication:

$$E_{\mu\nu}E_{\mu\nu}=-1, \tag{2·341}$$

$$E_{\mu\nu}E_{\mu\sigma}=-E_{\mu\sigma}E_{\mu\nu}=E_{\nu\sigma}, \tag{2·342}$$

$$E_{\mu\nu}E_{\sigma\tau}=\ \ E_{\sigma\tau}E_{\mu\nu}=iE_{\lambda\rho}, \tag{2·343}$$

where $\mu, \nu, \sigma, \tau, \lambda, \rho$ is any even permutation of 0, 1, 2, 3, 4, 5. For an odd permutation $E_{\mu\nu}E_{\sigma\tau}=-iE_{\lambda\rho}$.

The summation convention is not used in the above formulae. Unless otherwise stated we shall limit the summation convention to the row-and-column suffixes of matrices and wave vectors.§ In later developments the symbols $E_{\mu\nu}$ will be identified with matrices; they will then have the form

† Some writers use the term "complete set" for the group of linear expressions (in this case the E-numbers). The expressions in (2·25) would then be called "generators" of the complete set.

‡ *Amer. Journ. Math.* **1**, 350 (1878).

§ The summation convention is also employed when well-known formulae are quoted from general relativity theory.

$(E_{\mu\nu})_{\alpha\beta}$, where α and β indicate the element in the αth column and βth row of the matrix. In that case the summation convention will apply to α and β, but not to the suffixes μ, ν which distinguish one matrix of the set from another.

By (2·342) and (2·343) the E-symbols commute or anticommute according as they have no suffix or one suffix in common. We therefore obtain a sub-set of mutually anticommuting symbols by fixing one of the two suffixes and letting the other vary, e.g.

$$E_{30}, E_{31}, E_{32}, E_{34}, E_{35}.$$

We call such a sub-set a *pentad*. There are six different pentads; and each symbol is a member of two pentads. Our original symbols E_1, E_2, E_3, E_4, E_5 constitute the pentad with fixed suffix 0. It will be seen that, if we start from any of the other pentads and follow the same treatment, we reach the same complete set.

There exist also *triads*, i.e. sets of three mutually anticommuting E-symbols, which do not form parts of pentads, viz.

$$E_{\mu\nu}, E_{\nu\sigma}, E_{\sigma\mu} \quad (\mu \neq \nu \neq \sigma). \tag{2·35}$$

The maximum number of mutually commuting E-symbols (excluding i) is three; for no two of them can have a suffix in common, and therefore three symbols exhaust the six possible suffixes. The three commuting symbols are accordingly

$$E_{\mu\nu}, E_{\sigma\tau}, E_{\lambda\rho} \quad (\mu, \nu, \sigma, \tau, \lambda, \rho, \text{all different}). \tag{2·36}$$

We call such a set an *anti-triad*. Adding to it the symbol i, which commutes with all symbols, we obtain an *anti-tetrad*.

The two triads $$E_{\mu\nu}, E_{\nu\sigma}, E_{\sigma\mu}; \quad E_{\tau\lambda}, E_{\lambda\rho}, E_{\rho\tau}, \tag{2·37}$$

where μ, ν, σ, τ, λ, ρ are all different, are called *conjugate triads*. They have the property that every member of one triad commutes with every member of the other (see § 3·8).

We shall often employ an alternative single-suffix notation for the E-symbols (2·33), viz. E_μ ($\mu = 1, 2, \ldots 16$). It is then understood that the first five symbols form a pentad, and that

$$E_{16} = i, \tag{2·38}$$

but the order of the others is left unspecified. A general E-number is then

$$T = \sum_{\mu=1}^{\mu=16} t_\mu E_\mu, \tag{2·39}$$

the coefficients t_μ being algebraic numbers, real or complex. The individual terms $t_\mu E_\mu$ are called *components* of the E-number. The algebraic component $t_{16}E_{16}$, or it_{16}, will be called the *quarterspur* (abbreviated as qs). We have therefore

$$\text{qs}\, T = it_{16}. \tag{2·395}$$

If qs $T = 0$, T is said to be *degenerate*.

2·4. Linear Independence of the E-symbols.

(a) If a complete set is multiplied through by any one of its members, we obtain the same set in a different order, apart from algebraic factors ± 1 or $\pm i$. This follows from (2·34).

(b) If an E-number vanishes, every component is zero. For suppose that the E-number

$$t_a E_a + t_b E_b + \ldots + t_m E_m = 0, \qquad (2{\cdot}41)$$

the coefficients being non-zero. Multiply through by E_m; it follows from (a) that we obtain an expression of the same form and with the same number of terms as (2·41). The last term is $t_m E_m{}^2 = -t_m$ (or $it_m E_{16}$). Accordingly, let the result be

$$t_\alpha E_\alpha + t_\beta E_\beta + \ldots - t_m = 0. \qquad (2{\cdot}42)$$

Let E_τ be one of the symbols which anticommute with E_α. Multiply (2·42) firstly by initial E_τ and secondly by final E_τ, and add. Then

$$t_\alpha (E_\tau E_\alpha + E_\alpha E_\tau) + t_\beta (E_\tau E_\beta + E_\beta E_\tau) + \ldots - 2t_m E_\tau = 0. \qquad (2{\cdot}43)$$

The first term vanishes, and possibly some of the other terms; but the equation cannot wholly disappear since the last term does not vanish. If E_τ commutes with E_β, $(E_\tau E_\beta + E_\beta E_\tau) = \pm 2iE_\sigma$, where E_σ is another symbol of the set, by (2·343). Hence (2·43) reduces to an expression of the same form as (2·41) *but with fewer terms.*

By repeating the whole process as often as required we remove all terms except the last; we are then left with an equation containing just one non-zero term—which is absurd. Thus an equation of the form (2·41) is impossible unless all the coefficients are zero.

This shows that the E-symbols are not connected by any linear algebraic identity. In other words the set is complete but not redundant.

2·5. Miscellaneous Properties.

The following easily established properties of E-symbols are collected here for reference:

(a) Each symbol (except E_{16}) anticommutes with eight symbols, viz. the remaining members of the two pentads to which it belongs. It commutes with the remaining eight symbols, which include itself and E_{16}.

(b) Each symbol (except E_{16}) anticommutes with at least one member of any given tetrad. (A tetrad is formed by four members of a pentad.) For if the tetrad is $E_{01}, E_{02}, E_{03}, E_{04}$, the symbol $E_{\sigma\tau}$ has one suffix in common with one of these unless both σ and τ are 5. But σ and τ cannot be the same.

(c) If an E-number commutes with E_μ, every non-vanishing component commutes with E_μ. For the condition that $\Sigma t_\nu E_\nu$ commutes with E_μ is

$$\Sigma_\nu t_\nu (E_\nu E_\mu - E_\mu E_\nu) = 0.$$

Terms for which E_ν, E_μ commute disappear; terms for which E_ν, E_μ anticommute reduce to the form $\pm 2t_\nu E_\sigma$ by (2·342). No two terms reduce to

the same E_σ. By §2·4 (*b*) the coefficients of these surviving terms vanish separately; that is to say, $t_\nu = 0$ for those components E_ν which do not commute with E_μ.

(*d*) Similarly if an E-number anticommutes with E_μ, every non-vanishing component anticommutes with E_μ.

(*e*) If an E-number commutes with each member of a tetrad, it is an algebraic number. For by (*c*) its non-vanishing components commute with each member of a tetrad, and by (*b*) no E-symbol other than E_{16} can do this. The E-number therefore reduces to $E_{16}t_{16}$, or it_{16}.

(*f*) For any E-number T we have

$$\sum_{\mu=1}^{\mu=16} E_\mu T E_\mu = -16\,\mathrm{qs}\,T. \tag{2·51}$$

Consider the component $t_\nu E_\nu$. We have

$$E_\mu E_\nu E_\mu = -E_\nu, \quad \text{if } E_\nu, E_\mu \text{ commute},$$
$$= +E_\nu, \quad \text{if } E_\nu, E_\mu \text{ anticommute}.$$

Hence, if $\nu \neq 16$, we have by (*a*)

$$\Sigma_\mu E_\mu E_\nu E_\mu = -8E_\nu + 8E_\nu = 0.$$

For $\nu = 16$, $$\Sigma_\mu E_\mu E_{16} E_\mu = -16E_{16}.$$

Hence $$\Sigma_\mu E_\mu T E_\mu = -16E_{16}t_{16} = -16\,\mathrm{qs}\,T.$$

(*g*) The coefficients t_μ of an E-number satisfy

$$t_\mu = -\mathrm{qs}\,(TE_\mu) = -\mathrm{qs}\,(E_\mu T). \tag{2·52}$$

Let $S = TE_\mu$. Each component of S corresponds to a single component of T by § 2·4 (*a*). The quarterspur of S corresponds to the component $t_\mu E_\mu$ of T, and is therefore equal to $t_\mu E_\mu . E_\mu = -t_\mu$, which proves the theorem.

Combining (2·51) and (2·52) we obtain

$$16t_\sigma = -16\,\mathrm{qs}\,(TE_\sigma) = \Sigma_\mu E_\mu T E_\sigma E_\mu = \Sigma_\mu E_\mu E_\sigma T E_\mu. \tag{2·53}$$

(*h*) If S and T are E-numbers

$$\mathrm{qs}\,(ST) = -\Sigma_\mu s_\mu t_\mu = \mathrm{qs}\,(TS). \tag{2·54}$$

2·6. Reciprocals.

Let S and T be E-numbers. Generally there exist two quotients T/S, which are the E-numbers R, R' defined respectively by

$$RS = T, \quad SR' = T. \tag{2·61}$$

Considering the first of these equations, the vanishing of the E-number $RS - T$ requires that every component of it should vanish. We have therefore 16 equations (linear in r_μ) to determine the 16 coefficients r_μ of the E-number R. A solution will exist unless the determinant of the coefficients

of the r_μ vanishes. Since the coefficients of the r_μ are furnished by S, the existence or non-existence of a solution depends on S but not on T (assuming $T \neq 0$).

If there is no solution, i.e. if S fails as a divisor, S is said to be *singular*.

In particular, taking $T = 1$, an E-number S will have a reciprocal R (such that $RS = 1$) unless it is singular. A singular E-number has no reciprocal.

If $$RS = 1, \quad SR' = 1, \tag{2·62}$$

we have $$R = R(SR') = (RS)R' = R', \tag{2·63}$$

so that the same reciprocal is obtained by either definition. An E-number commutes with its reciprocal.

Let T be an E-number which commutes with S. Denoting the reciprocal of S by S^{-1}, we have $S^{-1}.TS.S^{-1} = S^{-1}.ST.S^{-1}$,

whence $$S^{-1}T = TS^{-1}.$$

So that an E-number which commutes with S commutes also with its reciprocal. We can also show that, if two E-numbers commute, their reciprocals (if any) commute.

If S is singular, the vanishing of the determinant of the coefficients makes it possible to obtain an infinitude of solutions of

$$RS = 0. \tag{2·64}$$

Any such solution R is called a *pseudo-reciprocal* of S. A pseudo-reciprocal is necessarily singular. If R is a pseudo-reciprocal of S, XR is also a pseudo-reciprocal of S, X being any E-number; for if $RS = 0$, $XRS = 0$. A product of E-numbers is singular if any of its factors are singular.

It is important to notice that, when S and T are E-numbers, the equation $ST = 0$ does not imply that either $S = 0$ or $T = 0$. There is an alternative, viz. that S and T are singular. If, however,

$$SE_\mu T = 0 \tag{2·65}$$

for *every* symbol E_μ of the complete set, then either $S = 0$ or $T = 0$. For suppose that $S \neq 0$. Then by (2·53)

$$16St_\sigma = \sum_\mu SE_\mu TE_\sigma E_\mu$$
$$= 0, \text{ by (2·65)}.$$

Hence $t_\sigma = 0$. Since this applies to every component t_σ, it follows that $T = 0$.

If S is singular and RS contains no non-algebraic terms, then

$$RS = 0. \tag{2·66}$$

For if RS were a non-vanishing algebraic quantity α, R/α would be the reciprocal of S, so that S could not be singular.

Since $-E_\mu.E_\mu = 1$, the symbols E_μ are not singular. Hence if $T \neq 0$, $E_\mu T \neq 0$.

2·7. Transformation of Complete Orthogonal Sets.†

Let $$F_\mu = qE_\mu q', \tag{2·71}$$

where $$qq' = q'q = 1. \tag{2·72}$$

Then the F_μ form a complete set having the same structure as the set of E_μ.

This is proved by showing that the relations (2·34) pass over unchanged from the E_μ to the F_μ. Taking, for example, (2·342)

$$\begin{aligned} F_{\mu\nu}F_{\mu\sigma} &= qE_{\mu\nu}q' \cdot qE_{\mu\sigma}q' \\ &= qE_{\mu\nu}E_{\mu\sigma}q' \quad \text{by (2·72)} \\ &= qE_{\nu\sigma}q' = F_{\nu\sigma}. \end{aligned}$$

Here q and q' may be E-numbers or they may involve entirely new symbols. We make no assumption as to their nature.

The converse theorem is that if E_μ, F_μ are two complete orthogonal sets, arranged in corresponding order so that F_μ, F_ν commute or anticommute according as E_μ, E_ν commute or anticommute, there exists a transformation (2·71) connecting them. We shall prove this under the restriction that

(*a*) The F_μ are E-numbers, or

(*b*) The F_μ are new symbols which commute with all the E_μ.

Let $$P = \alpha \sum_1^{16} F_\mu E_\mu, \quad P' = \alpha \sum_1^{16} E_\mu F_\mu, \tag{2·73}$$

α being an algebraic number. Consider the expression $F_\nu P E_\nu$. It has 16 terms of the form

$$\alpha F_\nu (F_\mu E_\mu) E_\nu.$$

If E_ν, E_μ anticommute, and therefore F_ν, F_μ anticommute, this becomes $-\alpha F_\mu F_\nu E_\mu E_\nu$, which is of the form $-\alpha F_\sigma E_\sigma$ by (2·342). If E_ν, E_μ and F_ν, F_μ commute, it becomes $\alpha F_\mu F_\nu E_\mu E_\nu$, which is of the form $\alpha (iF_\sigma)(iE_\sigma)$ or $-\alpha F_\sigma E_\sigma$ by (2·343). Thus $F_\nu P E_\nu$ gives the 16 terms of $-P$ in a different order; hence

$$F_\nu P E_\nu = -P. \tag{2·741}$$

Similarly $$E_\nu P' F_\nu = -P'. \tag{2·742}$$

Multiplying by final E_ν and final F_ν, respectively, these give

$$F_\nu P = P E_\nu, \quad E_\nu P' = P' F_\nu. \tag{2·75}$$

Case (*a*). F_μ is an E-number.

Multiplying together the two equations in (2·75), we have

$$E_\nu P' P E_\nu = P' F_\nu F_\nu P = -P'P.$$

Hence, multiplying initially by $-E_\nu$,

$$P'PE_\nu = E_\nu P'P.$$

Therefore $P'P$ commutes with every E_ν, i.e. with every symbol in the calculus. It is therefore an algebraic number. Reserving the singular case

† This transformation was introduced by G. Temple, *Proc. Roy. Soc.* A, **127**, 342 (1930).

$P'P=0$, for consideration in § 2·8, we can choose α so as to make $P'P=1$. It follows that $PP'=1$; and by (2·75)

$$PE_\nu P'=PP'F_\nu=F_\nu,$$

so that P is the required transformation operator q in (2·71).

Case (*b*). F_μ commutes with the E_μ.

Then $P=P'$. If P is multiplied by E_1F_1, we obtain the same 16 terms in a different order, except that those which commute with E_1 (and therefore with F_1) acquire a factor i^2 by (2·343). Thus 8 terms are reversed in sign. In the product $P\times P$, each term of P occurs 16 times, 8 times with the original sign and 8 times with reversed sign, except that $E_{16}F_{16}$ occurs 16 times with reversed sign. Hence $P^2=\alpha^2(-16E_{16}F_{16})=16\alpha^2$.

Taking $\alpha=\frac{1}{4}$, we have $PP'=P^2=1$. Then by (2·75)

$$PE_\nu P'=PP'F_\nu=F_\nu.$$

Hence the required transformation operator is

$$q=q'=\tfrac{1}{4}\Sigma E_\mu F_\mu. \tag{2·76}$$

2·8. The Singular Case.

In Case (*a*), but not in Case (*b*), it may happen that $PP'=0$ and the foregoing method of determining q breaks down. We shall show that nevertheless there is a transformation $F_\mu=qE_\mu q'$; but instead of $q=P$, we have $q=P^{(\sigma)}$, where

$$P^{(\sigma)}=\alpha\Sigma_\mu F_\mu E_\sigma E_\mu, \quad P'^{(\sigma)}=\alpha\Sigma_\mu E_\mu E_\sigma F_\mu, \tag{2·81}$$

and E_σ is one of E-symbols. The transformation previously given corresponds to $\sigma=16$; if that fails we try another value of σ, until we find one such that $P^{(\sigma)}P'^{(\sigma)}\neq 0$.

Ordinarily the vanishing of PP' does not imply that either P or P' is zero; but in the present case we have, by (2·75),

$$PE_\nu P'=PP'F_\nu=0$$

for *every* E_ν. Hence by (2·65) either $P=0$ or $P'=0$.

Let $$E_\mu^{(\sigma)}=(iE_\sigma)\,E_\mu\,(iE_\sigma)=-E_\sigma E_\mu E_\sigma. \tag{2·82}$$

Since this is a transformation of the form $qE_\mu q'$ with $qq'=1$, the $E_\mu^{(\sigma)}$ form a complete set. By (2·82) $E_\mu^{(\sigma)}=E_\mu$ or $-E_\mu$ according as E_σ commutes or anticommutes with E_μ; thus the transformation simply reverses the signs of eight members of the set. We call the set $E_\mu^{(\sigma)}$ a *reflection* of the set E_μ. Including the original set (reproduced when $\sigma=16$) there are sixteen different reflections, which correspond to the sixteen possible combinations of sign in an initial tetrad $\pm E_1$, $\pm E_2$, $\pm E_3$, $\pm E_4$.

By (2·51) $$\Sigma_\sigma E_\mu^{(\sigma)}=16\,\mathrm{qs}\,E_\mu. \tag{2·83}$$

Hence by (2·81)

$$\Sigma_\sigma P^{(\sigma)}E_\sigma = -16\alpha\Sigma_\mu F_\mu \operatorname{qs} E_\mu = -16\alpha F_{16}E_{16} = 16\alpha, \qquad (2\cdot84)$$

since $\operatorname{qs} E_\mu = 0$ unless $\mu = 16$. Hence at least one of the quantities $P^{(\sigma)}E_\sigma$ has a non-vanishing quarterspur. Therefore at least one of the quantities $P^{(\sigma)}$ does not vanish.

The quarterspurs of $P^{(\sigma)}E_\sigma$ and $E_\sigma P'^{(\sigma)}$ are equal; for they are the quarterspurs of $-\alpha\Sigma F_\mu E_\mu^{(\sigma)}$ and $-\alpha\Sigma E_\mu^{(\sigma)}F_\mu$, which are equal by (2·54). Hence, in securing that $P^{(\sigma)} \neq 0$, we also secure that $P'^{(\sigma)} \neq 0$.

Having found non-vanishing $P^{(\sigma)}$, $P'^{(\sigma)}$, we use these instead of P, P', and repeat our previous analysis as far as (2·75), obtaining

$$F_\nu P^{(\sigma)} = P^{(\sigma)}E_\nu, \quad E_\nu P'^{(\sigma)} = P'^{(\sigma)}F_\nu. \qquad (2\cdot85)$$

It will be found that E_σ remains passive in the middle of the expressions $F_\nu P^{(\sigma)}E_\nu$ and does not affect the argument. Proceeding to Case (*a*), we find as before that $P'^{(\sigma)}P^{(\sigma)}$ is algebraic. Further, it cannot vanish; for, as shown for P, P' at the beginning of this section, its vanishing would require that either $P^{(\sigma)}$ or $P'^{(\sigma)}$ is zero.

It may seem curious that we should be able to choose σ arbitrarily in the transformation $F_\mu = P^{(\sigma)}E_\mu P'^{(\sigma)}$. The explanation is that by changing σ we introduce a purely algebraic factor which is absorbed in α. Evidently the transformation could be further generalised by substituting an arbitrary E-number in place of E_σ.

We notice for future reference that there is at least one reflection $E_\mu^{(\sigma)}$ which is connected with F_μ by the unmodified transformation $q = \alpha\Sigma F_\mu E_\mu^{(\sigma)}$, $q' = \alpha\Sigma E_\mu^{(\sigma)}F_\mu$. For, choosing σ so that $P^{(\sigma)}$, $P'^{(\sigma)} \neq 0$, we have

$$\alpha\Sigma_\mu F_\mu E_\mu^{(\sigma)} = -\alpha\Sigma F_\mu E_\sigma E_\mu E_\sigma = -P^{(\sigma)}E_\sigma \neq 0,$$

$$\alpha\Sigma_\mu E_\mu^{(\sigma)}F_\mu = -\alpha\Sigma E_\sigma E_\mu E_\sigma F_\mu = -E_\sigma P'^{(\sigma)} \neq 0.$$

2·9. Application to Relativity.

The transformation $F_\mu = qE_\mu q'$ is formally the same as the transformation (1·463) of a mixed tensor $T' = qTq'$. Limiting ourselves for the present to Case (*a*), q is now a non-singular E-number instead of a non-singular matrix. We shall find in Chapter III that fourfold matrices are a special representation of E-numbers. Thus it is appropriate to generalise the definitions of tensors in Chapter I by substituting symbolic E-numbers for matrices.

In physical applications we shall call a complete set of E_μ a *symbolic frame*. By the above transformation we obtain a different but equivalent symbolic frame F_μ. If the E_μ are taken to be mixed wave tensors, the change from one symbolic frame to an equivalent frame is a tensor transformation. A change of symbolic frame is then part of a general change of system of description, and other quantities occurring in the description are

changed simultaneously according to the wave tensor character assigned to them in the group of descriptions contemplated.

The present analytical theory is being developed to serve as a tool in physical investigations, and we cannot pledge ourselves always to use the tool in one particular way. But the primary application will be that certain characteristics of a physical system are described by E-numbers which are invariant for changes of the system of description, and therefore correspond in conception to an absolute structure transcending our variable description. Let $T = \Sigma t_\mu E_\mu$ be one of these invariant E-numbers. The invariance requires that when we transform to a new symbolic frame E_μ', the coefficients t_μ are changed to t_μ' so that

$$T = \Sigma t_\mu E_\mu = \Sigma t_\mu' E_\mu'. \tag{2·91}$$

The arrays t_μ, t_μ' are regarded as components of the same physical entity referred to two different reference frames E_μ, E_μ'.

It is convenient to call the transformation $E_\mu \to E_\mu'$ a *rotation* of the symbolic frame. Rotation is here given a somewhat generalised meaning; the emphasis is on the fact that it is a type of change which does not involve any intrinsic distortion of the frame. The frame E_μ' has the same intrinsic structure as the frame E_μ, namely that expressed by equations (2·34).

Having defined rotations of the frame we can now define corresponding rotations of a physical system described by E-numbers. Consider a system described by invariant E-numbers $T = \Sigma t_\mu E_\mu$, $U = \Sigma u_\mu E_\mu$, etc. Let the system undergo a change such that $T \to T'$, $U \to U'$, etc., where

$$T' = \Sigma t_\mu E_\mu', \quad U' = \Sigma u_\mu E_\mu'. \tag{2·92}$$

Then the new physical system is constructed in the frame E_μ' according to the same specification as that by which the original system was constructed in the frame E_μ. In other words the system has rotated with the frame. Clearly the systems (T, U, ...) and (T', U', ...) have the same kind of equivalence as the frames. They are intrinsically similar, as the frames are intrinsically similar.

Normally the rotation of a physical system is described by referring it to a fixed frame. We therefore require the components of T' in the original frame E_μ. Denoting these components by t_μ' (not the same t_μ' as in (2·91)) the condition is

$$T' = \Sigma t_\mu E_\mu' = \Sigma t_\mu' E_\mu. \tag{2·93}$$

The transformation $t_\mu \to t_\mu'$ represents a rotation of the physical system relative to the fixed frame E_μ. Since $E_\mu' = q E_\mu q'$, we have by (2·93)

$$\Sigma t_\mu' E_\mu = q\,(\Sigma t_\mu E_\mu)\, q'. \tag{2·94}$$

The nature of the transformations of t_μ determined by (2·94) will be studied in detail in Chapter IV.

Any change $t_\mu \to t_\mu'$ represents some imaginable change of the physical system described by t_μ. The peculiarity of the transformations which satisfy

(2·94) is that the new system is intrinsically similar to the old, and the change is therefore pictured geometrically as a rotation without distortion. More generally we call such a change a *relativity transformation*. It can be detected (if at all) by observing relations to extraneous physical objects that do not form part of the system to which the relativity transformation is applied.

The relativity of our orthogonal symbolic frames is precisely analogous to the relativity of Galilean frames of space and time. Space-time frames are all alike initially. If we speak of a frame x, y, z, t, it is impossible to define in an absolute way which frame, out of an infinite number of equivalent frames, we refer to. But when once we have selected and labelled an initial frame A, any other frame B can be defined relatively to it by specifying the space rotation and Lorentz transformation which would convert A into B. Similarly we cannot define in an absolute way the frame E_μ which we select initially. But any other frame E_μ' can be defined relatively to E_μ by specifying the transformation symbol q (which is an E-number of the form $\Sigma q_\mu E_\mu$) connecting E_μ and E_μ'. We use these symbolic frames as the basis of a relativity theory which (we shall find) includes, but is somewhat more comprehensive than, the relativity of Galilean frames of space and time.

Attention may be called to the perfect adaptation of the mathematical symbolism to the physical conditions. Owing to relativity we are unable to define in an absolute way the physical frame initially selected, which we label E_μ. It is therefore appropriate that we should be equally unable to define in an absolute way the label E_μ which we affix. For the set of symbols E_μ is only defined by its structural properties (2·34), and these apply equally to E_μ' or to any other complete orthogonal set. The complete physical equivalence is therefore represented by a complete mathematical equivalence. We lose this perfect adaptation when we use special kinds of E_μ, e.g. matrices.

There is no absolute distinction between a rotation of the physical system and a rotation of the frame in the opposite direction; and in elementary theory the term "relativity rotation" is applied indifferently to rotations of the physical system and of the frame. But after the first results of this equivalence have been gleaned, there is seldom anything to be learned by introducing rotations of the frame. If the frame is rotated, we have to transform simultaneously the specification of all objects, fields, boundary conditions (including boundary conditions at "infinity"), normalising conditions, etc., concerned in the problem. On the other hand, keeping the frame fixed, we can introduce relativity rotations of a particular object, leaving the other objects concerned in the system unchanged. In order that it may possess independent relativity rotations, the object must be conceived as separable from the rest of the system contemplated. A separable

object will have a structure described by invariant E-numbers T, U, ..., so that transformations of the form (2·94) represent displacements of the object without intrinsic change of its structure.

Thus in later developments we are concerned with independent relativity rotations of individual objects, which provide much wider scope for the application of relativistic principles than a rotation of their common frame.† For this reason, "relativity transformation" will normally mean displacement without intrinsic change of an object referred to a fixed frame, though it may also be applied to a rotation of the frame if occasion arises.

In practice an object cannot be rigorously separated from its surroundings; if it could be separated, it would not be accessible to observation. But that does not do away with the usefulness of the conception of a separable object. One of the greatest achievements of current quantum theory is that it has found a rigorous method of avoiding this dilemma. An incompletely separated *object* is represented as a probability distribution over completely separated *states*. The environment then affects, not the state, but the probability attached to the state. We may therefore, with all rigour, apply relativity transformations to the E-numbers describing the states—provided, of course, that the states are such as can be specified by invariant E-numbers. This last reservation is liable to be overlooked; and E-numbers (or the equivalent Dirac wave functions) have often been applied to states which obviously do not possess the relativistic properties which E-numbers are designed to represent.

From one point of view the assumption that there exist in nature equivalent 16-fold frames, which can therefore be appropriately represented by equivalent sets of E-symbols, is a hypothesis—the fundamental hypothesis of our theory. But actually we appeal to an epistemological principle which goes deeper than that. We will call it the "Principle of the Blank Sheet".

Physics is concerned with the problem of distinguishing and classifying the distinctions of objects, states, events. Exact measurement is a process of determining and classifying minute distinctions. To develop a theory of the characteristics which can be distinguished and of the measurement of the distinction, we require a blank sheet to write on—not a sheet already scribbled over with vaguely recognised distinctions. A group of intrinsically indistinguishable frames is chosen as the basis of a description of the universe, in order that the theory of distinguishable or measurable phenomena to be erected on it may go down to the very origin of their distinction. In

† The above remarks refer to the more usual problems of quantum theory in which a number of objects are referred to a single frame. An important intermediate step is the consideration of the relativity rotations of two objects referred to a *double* frame $E_\mu F_\nu$. In this case any combination of rotations of the two objects is equivalent to a transformation to an equivalent double frame $E_\mu' F_\nu'$. This (rather unusual) development is of great importance in the special problems treated in the present book.

practice we *do* distinguish frames of space-time otherwise than by their transformation relations to one another; we distinguish them as being at rest relatively to the earth, sun, etc., or as having a special orientation with respect to the earth's gravitational field; but we conceive frames of space-time as initially indistinguishable in order that these distinctions may be properly inserted in the development of the theory and not hidden in its initial assumptions. We do *not* assert that there exist intrinsically indistinguishable 16-fold frames of reference in the physical world; it is only for an ideally simplified universe that this would be true. Our principle is that such distinguishability of the frames as occurs must be treated as a positive characteristic to be represented by appropriate symbols and combined in a unified theory with the other distinctions studied in physics. To exhibit a positive characteristic, we have to imagine a frame which initially lacks it.

Our mode of thought requires us to formulate some kind of frame or background for physical phenomena. Let the background be a white sheet to show up the phenomena, not a jazz-painted camouflage against which they may lie undetected.

Case (*b*), in which the equivalent frames E_μ, F_μ consist of entirely different symbols, has also an important physical application. Since q is now a mixture of E-symbols and F-symbols, the relation between the two frames is not describable by reference to the E-frame only. If we have three such frames E, F, G, the transformations q_{EF}, q_{EG} cannot be compared, and there is no meaning in saying that the change from E to F is greater or less than the change from E to G. Equivalent frames of this kind are required when we deal with the properties of two similar atoms or two electrons, conceived as non-interacting. Two electrons are intrinsically similar (or equivalent) but are not the same; we cannot specify different degrees or different kinds of not-the-sameness, as we do for equivalent space-time frames. If there is interaction the case is somewhat altered, and the electrons are not so definitely distinct; but here again the Principle of the Blank Sheet requires us to start with frames corresponding to non-interaction, into which interaction is introduced as an explicit perturbation.

CHAPTER III

THE RESOLUTION OF MATRICES

3·1. Four-point Matrices.

We are now going to show that fourfold matrices may be expressed as E-numbers; so that the theory developed in Chapter II has a particular application to matrices.

First consider the six matrices

$$S_\alpha = \begin{matrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{matrix} \qquad S_\beta = \begin{matrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix} \qquad S_\gamma = \begin{matrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{matrix}$$

$$D_\alpha = \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix} \qquad D_\beta = \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix} \qquad D_\gamma = \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

The system of nomenclature is that the suffixes α, β, γ have reference to the three ways of pairing four numbers, viz. 12, 34; 13, 24; 14, 23.

We also introduce two alternative notations for the unit matrix, viz.

$$S_\delta = D_\delta = 1.$$

The following results of matrix multiplication are easily verified:

$$S_\alpha S_\beta = S_\gamma, \qquad D_\alpha D_\beta = D_\gamma; \tag{3·111}$$

$$S_\alpha^2 = 1, \qquad D_\alpha^2 = 1; \tag{3·112}$$

$$S_\alpha D_\alpha = D_\alpha S_\alpha, \quad S_\alpha D_\beta = -D_\beta S_\alpha; \tag{3·113}$$

with similar results obtained by permuting α, β, γ, but not δ.

The commutative properties may be summarised as follows $(a, b = \alpha, \beta, \gamma, \delta)$:

$$S_a S_b = S_b S_a, \quad D_a D_b = D_b D_a, \quad S_a D_b = (ab)\, D_b S_a, \tag{3·12}$$

where

$$\left.\begin{aligned} (ab) = (ba) = 1 \quad & \text{if } a = \delta, \text{ or } b = \delta, \text{ or } a = b \\ = -1 \quad & \text{otherwise} \end{aligned}\right\}. \tag{3·13}$$

The product of any number of these matrices in any order can be reduced to one of the sixteen forms:

$$\pm S_a D_b \quad (a, b = \alpha, \beta, \gamma, \delta).$$

For we can bring all the S's to the beginning and the D's to the end by applying (3·12), and then reduce the S's to a single S and the D's to a single D by applying (3·112) and (3·111). If either the S factor or the D factor disappears, we insert the unit matrix S_δ or D_δ to preserve homogeneity.

Thus the sixteen forms constitute a complete set in the sense explained in § 2·2. If we call a linear function of them with algebraic coefficients an *SD*-number, the operations of addition, subtraction and multiplication applied to *SD*-numbers will always yield *SD*-numbers.

3·2. Pentads.

It follows from (3·12) that

$$(S_a D_b)(S_c D_d) = (bc)(ad)(S_c D_d)(S_a D_b). \tag{3·21}$$

Hence the condition that $S_a D_b$ and $S_c D_d$ anticommute is

$$(bc)(ad) = -1. \tag{3·22}$$

Let us write down the matrices $S_c D_d$ which anticommute with $S_\alpha D_\delta$. Here $b = \delta$, so that $(bc) = 1$. Hence $(ad) = -1$; and since $a = \alpha$, we have $d = \beta$ or γ. The suffix c can have any value. Hence the matrices are

$$S_\alpha D_\beta,\ S_\beta D_\beta,\ S_\gamma D_\beta,\ S_\delta D_\beta,\ S_\alpha D_\gamma,\ S_\beta D_\gamma,\ S_\gamma D_\gamma,\ S_\delta D_\gamma. \tag{3·23}$$

Selecting one of these, $S_\delta D_\beta$, we find in a similar way that the following anticommute with it:

$$S_\alpha D_\alpha,\ S_\alpha D_\beta,\ S_\alpha D_\gamma,\ S_\alpha D_\delta,\ S_\gamma D_\alpha,\ S_\gamma D_\beta,\ S_\gamma D_\gamma,\ S_\gamma D_\delta. \tag{3·24}$$

Hence the following anticommute both with $S_\alpha D_\delta$ and $S_\delta D_\beta$:

$$S_\alpha D_\beta,\ S_\alpha D_\gamma,\ S_\gamma D_\beta,\ S_\gamma D_\gamma. \tag{3·25}$$

The first of these is the product of $S_\alpha D_\delta$ and $S_\delta D_\beta$. A symbol which anticommutes with two symbols necessarily commutes with their product; thus no further matrices can anticommute with the triad:

$$S_\alpha D_\delta,\ S_\delta D_\beta,\ S_\alpha D_\beta. \tag{3·26}$$

It will be found that the remaining three matrices in (3·25) anticommute with each other, so that

$$S_\alpha D_\delta,\ S_\delta D_\beta,\ S_\alpha D_\gamma,\ S_\gamma D_\beta,\ S_\gamma D_\gamma$$

constitute a pentad of mutually anticommuting matrices.

Dropping the superfluous S_δ and D_δ, and inserting a factor i where necessary to make the square of the matrix equal to -1, the pentad is

$$iS_\alpha,\ iD_\beta,\ iS_\gamma D_\gamma,\ S_\alpha D_\gamma,\ S_\gamma D_\beta. \tag{3·27}$$

These five matrices accordingly satisfy the same conditions (2·23) as a pentad of E-symbols, and constitute a particular identification of E_1, E_2, E_3, E_4, E_5. If identified in this order they are found to satisfy (2·22). All the theorems of Chapter II then have an application to matrices.

The complete set E_μ can accordingly be identified with the complete set $i^{\frac{1}{2}(ab)} S_a D_b$, the factor i being inserted when $(ab) = +1$ in order to make the square equal to -1.

The five other pentads can be found from the theory of E-symbols, or more simply by permuting α, β, γ in (3·27).

A set of four anticommuting four-point matrices was first introduced into physical theory by P. A. M. Dirac in his wave equation of an electron. The particular matrices used by Dirac form part of one of the pentads here found. It was shown by J. v. Neumann that the complete set consisted of 16 matrices.† The complete set $S_a D_b$ was first studied in this connection by the author.‡ It has been pointed out by P. du Val that a similar analysis had been developed in connection with the theory of Kummer's Quartic Surface.§

3·3. Components of Matrices.

In § 3·2 we have found a representation of the E-symbols, and hence of all E-numbers, by fourfold matrices. We shall now prove the converse, viz. that every fourfold matrix will represent an E-number; that is to say, any fourfold matrix T can be expressed in the form

$$T = \sum_{1}^{16} t_\mu E_\mu, \tag{3·311}$$

where E_1, E_2, E_3, E_4, E_5 are the matrices (3·27).

The meaning of (3·311) will be clearer if we insert the row-and-column suffixes α, β of the matrices, viz.

$$T_{\alpha\beta} = \sum_{\mu} t_\mu (E_\mu)_{\alpha\beta}. \tag{3·312}$$

Considering in succession the 16 combinations of suffixes α, β, we have 16 equations to determine the 16 algebraic coefficients t_μ. The values of t_μ are unique; for if there were another set of values t_μ', we should have by subtraction

$$0 = \sum_{\mu} (t_\mu - t_\mu') E_\mu.$$

Hence by § 2·4 (*b*), $t_\mu - t_\mu' = 0$.

As we shall presently solve these equations for t_μ, it is not necessary to stop to prove here that the condition for the existence of a solution (non-vanishing of the determinant of the coefficients) is satisfied.

We call t_μ (or $E_\mu t_\mu$) a *component* of T, and μ a *matrix suffix*, as distinguished from the *elements* $T_{\alpha\beta}$ and the *row-and-column suffixes* α, β.

The sum of the diagonal elements of a matrix is called the *spur*, and will be denoted by $\{T\}$.

The matrices $S_a D_b$ have no diagonal elements unless $a = \delta$. Also we see from the definitions of D_α, D_β, D_γ in § 3·1 that their spurs vanish. Hence

† *Zeits. für Physik*, **48**, 881 (1928).

‡ *Proc. Roy. Soc.* A, **121**, 524 (1928).

§ The connection with the Kummer collineation group has been treated fully by O. Zariski, *Amer. Journ. Math.* **54**, 466 (1932).

$\{S_a D_b\}=0$, except $\{S_\delta D_\delta\}=4$. Changing to the E_μ notation, we have

$$\{E_\mu\}=0 \quad \text{for } \mu=1,2,\dots 15, \quad \{E_{16}\}=4i. \tag{3·32}$$

Taking the spur of (3·311), we have

$$\{T\}=\sum_\mu t_\mu\{E_\mu\}=4it_{16}=4\,\mathrm{qs}\,T \tag{3·33}$$

by (2·395). Thus the spur of a matrix is (appropriately) four times the quarterspur. It is to be remembered that for a general symbolic E-number (not identified with a matrix) the diagonal sum would have no meaning. It is for that reason that we have introduced the quarterspur as a more general characteristic, not implying matrix representation.

By (2·52) $$t_\mu=-\mathrm{qs}\,(E_\mu T)=-\tfrac{1}{4}\{E_\mu T\} \tag{3·34}$$

by (3·33). We have thus an explicit formula for the components t_μ of a matrix.

Owing to the great importance of (3·34) it may be desirable to give a direct proof. Multiply both sides of (3·311) by E_ν, and take the diagonal sum; we have

$$\{E_\nu T\}=\sum_\mu t_\mu\{E_\nu E_\mu\}.$$

Now $E_\nu E_\mu$ reduces to a single symbol, whose spur vanishes by (3·32) except when $\mu=\nu$. Hence

$$\{E_\nu T\}=t_\nu\{E_\nu^2\}=-t_\nu\{1\}=-4t_\nu,$$

which is equivalent to (3·34).

We can write (3·34) in a form which avoids the use of the symbol { }. Inserting row-and-column suffixes (and temporarily dropping the matrix suffix),

$$ET=E_\alpha{}^\beta T_\beta{}^\gamma, \quad \{ET\}=E_\alpha{}^\beta T_\beta{}^\alpha.$$

Divide the symbol T into two portions each carrying a suffix, thus,

$$T_\alpha{}^\beta=\mathrm{T}_\alpha\Gamma^\beta.$$

Since the suffixes are explicitly indicated, we may rearrange the order of the factors,

$$\{ET\}=E_\alpha{}^\beta\mathrm{T}_\beta\Gamma^\alpha=\Gamma^\alpha E_\alpha{}^\beta\mathrm{T}_\beta=\Gamma E\mathrm{T},$$

the suffixes being omitted after the rearrangement since they follow the chain rule. Thus (3·34) becomes

$$t_\mu=-\tfrac{1}{4}\Gamma E_\mu\mathrm{T}. \tag{3·35}$$

We call T and Γ *symbolic factors* of T.

In particular, if a matrix J is the outer product of two vectors ψ, χ^*, so that

$$J=\psi\chi^*, \tag{3·36}$$

the components are

$$j_\mu=-\tfrac{1}{4}\chi^* E_\mu\psi. \tag{3·37}$$

The *expectation value* of an operator X with respect to wave vectors ψ and χ^* is defined to be† $\bar{X} = \chi^* X \psi \div \chi^* \psi.$ (3·38)

It follows from (3·37) that the expectation value of E_μ is ij_μ/j_{16}.

3·4. General Orthogonal Frames.

The *SD* matrices which constitute the symbolic frame used in § 3·3 are of a special type, called *four-point matrices*, since they have only four non-vanishing elements. By § 2·7, Case (*a*), an equivalent frame E_μ' is obtained by the transformation $E_\mu' = qE_\mu q' \quad (qq' = 1),$ (3·41)

where q is any non-singular E-number, and therefore in the present application any non-singular fourfold matrix.

The matrices E_μ of the new frame will not generally be four-point matrices. We have therefore to consider whether the results of § 3·3 will apply to the new frame.

The spur $\{E_\mu\}$ is invariant for the transformation (3·41). For

$$\begin{aligned}\{E_\mu'\} &= (E_\mu')_{\alpha\alpha} = q_{\alpha\beta}(E_\mu)_{\beta\gamma} q_{\gamma\alpha}' = q_{\gamma\alpha}' q_{\alpha\beta}(E_\mu)_{\beta\gamma} \\ &= (q'qE_\mu)_{\gamma\gamma} = \{E_\mu\},\end{aligned} \tag{3·42}$$

since $q'q = 1$. Hence the formulae (3·32) apply equally to the frame E_μ'. Except in calculating the spur, no use was made of the special properties of *SD* matrices; and therefore all the results in § 3·3 apply to E_μ'.

In particular the formulae (3·35) and (3·37) for the components apply to any orthogonal frame of matrices.

Up to the end of Chapter VI we shall (unless otherwise stated) take the E_μ to be general fourfold matrices which satisfy the conditions for a complete set. We shall not specify the particular set of matrices used. This is in accordance with the relativity principle in § 2·9, that there can be no absolute description of the reference frame initially chosen; but if other frames are subsequently introduced they can be defined relatively to the first frame by stating the components (in the first frame) of the transformation matrix q. The frame $S_a D_b$ has served its purpose in enabling us to construct the whole set of equivalent frames; but it has no special significance in physics, since the structure of the commutation relations is common to all the frames. For example, we must not think of $S_a D_b$ as being physically distinguished from other legitimate frames in the way that Galilean coordinates are distinguished from other legitimate coordinates. Its apparent distinctiveness (shown in the simplicity of the matrices) is really a misfit between the

† This is a somewhat generalised definition of expectation value. In current theory the term is restricted to an expectation value with respect to *one* wave vector ψ; χ is then replaced in the formula by the complex conjugate of ψ. It must not be assumed that the familiar properties (e.g. that the expectation value is intermediate between the greatest and least eigenvalues) hold for the generalised definition.

physical structure and the mathematical expression of it by matrices. If we keep to general symbolic E-numbers no such misfit occurs, and in that respect they give a closer representation of the actualities of physics than the matrix representation does.

It may be asked, What do we gain by introducing matrices instead of general E-numbers? Ultimately I think we gain nothing. I do not think that there is anything in the physical constitution of the systems to which we apply this calculus that is represented in the matrices and unrepresented in the general E-numbers. The main justification for using a particular representation is that it simplifies the algebra in practical problems. Thus in Einstein's theory we introduce special coordinates for the discussion of the phenomena of the solar system, since the analysis of these phenomena would be intolerably difficult if we retained general coordinates throughout. We shall sometimes use the frame of four-point matrices in this way to establish results known to be invariant, which it is therefore sufficient to prove in any one frame of reference. On the other hand we are liable to lose valuable insight by premature introduction of special frames or special coordinates. Temple has shown that, even in so special a problem as the determination of the energy levels of the hydrogen atom, matrices are not required, and the work can be carried out with general E-symbols; his determination appears to me not only more illuminating but actually much simpler as regards algebraic calculation than the proofs previously given in terms of matrices (§ 9·3). In any case the use of matrix representation expressly for the purpose of facilitating calculation is a very different matter from its use in the formulation of the fundamental laws of physics.

But, whatever the ideal course, I am here limited by the fact that I do not propose to reinvestigate the whole quantum theory. I must develop the present relativity theory up to a point at which it meets the accepted results of quantum theory which are soundly (if unaesthetically) established. These results are given in matrix representation by Dirac and others, and the conventional nomenclature and definitions have reference to the matrix representation. I must have an eye on the theory that I am steering to meet before I actually make contact with it; therefore it seems unwise to postpone the transition to matrix representation for long. Meanwhile the knowledge that there is an equivalent theory in terms of general symbols is reassuring; for I cannot believe that anything so ugly as the multiplication of matrices is an essential part of the scheme of nature.

In § 2·7, Case (*b*), we may take the E_μ to be matrices and the F_μ to be general symbols, or *vice versa*; then q will be a mixture of matrices and general symbols. Thus the use of matrix representation does not entirely cut us off from general symbols; the gap can be bridged by an ordinary tensor transformation. I shall be talking chiefly about matrix frames; but if you will

inscribe q on the front cover of the book and q' on the back cover—then I am talking about general symbolic frames!

There are two kinds of property which at first sight seem to be expressed more simply in matrix calculus than in general symbolic calculus:

(1) A matrix can or cannot be resolved into two factors. We shall call a factorisable matrix a *pure* matrix. Thus we can recognise a distinction between pure and impure matrices, which is not apparent in the corresponding general symbols; the "factors" of a general E-number are an undefined conception. But purity of a matrix is an invariant property for all wave tensor transformations, since the two factors (vectors) transform separately. There is therefore some invariant characteristic of an E-number which corresponds to the factorisability of all its matrix representations. This characteristic is found to be idempotency (§ 5·6). Purity is expressed quite as easily by idempotency in symbolic calculus as by factorisability in matrix calculus.

(2) In the specimen pentad (3·27) three matrices are imaginary and two are real. This partition persists in all pentads (§ 3·5); and it is of great importance in physics, being the foundation of the distinction between space and time. To ascribe real or imaginary character to general symbols would involve something not expressible in terms of their commutability relations. It is to be remembered that the E_μ are all square roots of -1, whether they are represented by real or imaginary matrices. In the case of four-point matrices the imaginary matrices are symmetrical and the real matrices antisymmetrical (for interchange of rows and columns), and the distinction can be equivalently described by reference to symmetry; but the property of symmetry or antisymmetry is not invariant for tensor transformations. Matrix representation seems to afford the easiest way of expressing this distinction; but there would be no great difficulty in working out an alternative treatment by general symbolic methods if desired.

From Chapter VII onwards our point of view changes, and we shall generally restrict the E_μ to four-point matrices (SD matrices). That is because we have finished contemplating the "blank sheet" and are beginning to write something on it; and the property of symmetry or antisymmetry of the matrices of a certain frame is one of the first things that we write.

3·5. Real and Imaginary Matrices.

A matrix is said to be real if all its elements are real, and imaginary if all its elements are imaginary. If any of the elements are complex, or if some are real and some imaginary, the matrix is said to be complex. We shall show that, if complex matrices are excluded, three members of a pentad are imaginary and two are real.

We first prove by a *reductio ad absurdum* that five imaginary matrices

cannot form a pentad. Suppose then that F_1, F_2, F_3, F_4, F_5 are imaginary matrices forming a pentad; this is connected by a tensor transformation with the known pentad (3·27) containing three imaginary matrices E_1, E_2, E_3 and two real matrices E_4, E_5. By § 2·7 the transformation connecting complete sets E_μ, F_μ is

$$F_\mu = PE_\mu P', \quad P = \alpha \Sigma F_\mu E_\mu, \quad P' = \alpha \Sigma E_\mu F_\mu, \quad PP' = 1. \quad (3·51)$$

The singular case is avoided by using an appropriate reflection of the pentad (3·27); for, as shown at the end of § 2·8, there is at least one reflection which gives non-zero P and P'.

Write
$$P = R + iS, \quad P' = R' + iS', \quad (3·52)$$

where R, R', S, S' are real matrices. Then, since $P'P = 1$,

$$R'R - S'S = 1, \quad R'S + S'R = 0. \quad (3·53)$$

By (2·75)
$$F_1 P = PE_1, \quad E_1 P' = P'F_1.$$

Hence, separating the real and imaginary parts,

$$F_1 R = RE_1, \quad F_1 S = SE_1, \quad E_1 R' = R'F_1, \quad E_1 S' = S'F_1,$$

so that
$$E_1 R' R E_1 = R' F_1 F_1 R = -R'R.$$

Hence $E_1 R'R = R'RE_1$. Similarly $R'R$ commutes with E_2 and E_3. Therefore by § 2·5 (*c*) it consists of components which commute with E_1, E_2, E_3. This restricts it to the form
$$R'R = a + bE_{45}. \quad (3·54)$$

Since E_{45} is a real matrix, a and b are real coefficients.

Again, separating the real and imaginary parts of $F_4 P = PE_4, E_4 P' = P'F_4$, we have

$$F_4 R = iSE_4, \quad F_4 S = -iRE_4, \quad E_4 R' = iS'F_4, \quad E_4 S' = -iR'F_4,$$

so that
$$E_4 R' S E_4 = S' F_4 F_4 R = -S'R = R'S$$

by (3·53). Therefore $R'S$ anticommutes with E_4. We can show similarly that it anticommutes with E_{14}, E_{24}, E_{34}. This restricts it to the form

$$R'S = cE_{45} = -S'R, \quad (3·55)$$

where c is real.

By (3·54) and (3·55)

$$R'P = R'(R + iS) = a + bE_{45} + icE_{45},$$
$$P'R = (R' + iS')R = a + bE_{45} - icE_{45}.$$

Therefore

$$R'R = R'PP'R = (a + bE_{45})^2 + c^2 E_{45}^2 = a^2 - b^2 - c^2 + 2abE_{45}. \quad (3·56)$$

Comparing with (3·54), we have

$$a = \tfrac{1}{2}, \quad a^2 - b^2 - c^2 = a,$$

so that $b^2 + c^2 = -\tfrac{1}{4}$, which is impossible since b and c are real.

We can show similarly that a pentad of four real matrices F_1, F_2, F_4, F_5

and one imaginary matrix F_3 leads to a contradiction. For, denoting the real matrices in the standard pentad by E_1, E_2 and the imaginary matrices by E_3, E_4, E_5, we have again two matrices F_4, F_5 whose character (real or imaginary) is opposite to that of the corresponding matrices E_4, E_5, and the proof applies without alteration.

No other case arises, since by (2·22) the number of imaginary matrices in a pentad is necessarily odd. Thus the only possible partition of matrices in a pentad is three imaginary and two real.

The theorem has been generalised to matrices of m rows and columns by M. H. A. Newman.† If $m = 2^q p$, where p is odd, the maximum number of matrices in an anticommuting set is $2q+1$; and of these $q+1$ are imaginary and q real.

A case that might possibly be of physical interest is $m = 16$. The maximal anticommuting sets are then nonads with five imaginary and four real matrices. A nonad can be constructed as follows: The sixteen rows are designated by double suffixes $\alpha\beta$ (α, $\beta = 1, 2, 3, 4$). Then if E_μ denotes a 4-rowed matrix correlated to the first suffix, and F_μ the same matrix correlated to the second suffix, the outer product $E_\mu F_\nu$ is a 16-rowed matrix.‡ An example of a nonad is

$$iE_{23}F_1,\ iE_{31}F_1,\ iE_{12}F_1,\ iE_4F_2,\ iE_5F_2,\ iE_{45}F_2,\ F_3,\ F_4,\ F_5. \quad (3\cdot57)$$

It is constructed by means of a pair of conjugate triads of E matrices (see (3·82)).

3·6. Determinant of an E-number.

It is useful to have before us the explicit expression for the matrix which represents a general E-number $T = \Sigma t_\mu E_\mu$ with some standard identification of the matrices E_μ. The following is the matrix representing T, when the matrices E_1, E_2, E_3, E_4, E_5 are taken to be iS_α, iD_β, $iS_\gamma D_\gamma$, $S_\alpha D_\gamma$, $S_\gamma D_\beta$ as in (3·27). We write τ_μ for it_μ.

$$\begin{array}{llll}
\tau_{16}+\tau_{35}+\tau_2+\tau_{14}, & \tau_1+\tau_{42}+t_{21}+t_4, & \tau_{43}+t_{45}+\tau_{15}+t_{31}, & \tau_{52}+t_{23}+t_5+\tau_3,\\
\tau_1+\tau_{42}-t_{21}-t_4, & \tau_{16}+\tau_{35}-\tau_2-\tau_{14}, & \tau_{52}+t_{23}-t_5-\tau_3, & \tau_{43}+t_{45}-\tau_{15}-t_{31},\\
\tau_{43}-t_{45}+\tau_{15}-t_{31}, & \tau_{52}-t_{23}+t_5-\tau_3, & \tau_{16}-\tau_{35}+\tau_2-\tau_{14}, & \tau_1-\tau_{42}+t_{21}-t_4,\\
\tau_{52}-t_{23}-t_5+\tau_3, & \tau_{43}-t_{45}-\tau_{15}+t_{31}, & \tau_1-\tau_{42}-t_{21}+t_4, & \tau_{16}-\tau_{35}-\tau_2+\tau_{14}.
\end{array} \quad (3\cdot61)$$

The columns correspond to the first suffix and the rows to the second suffix of $T_{\alpha\beta}$.

The determinant formed by (3·61) can be evaluated. It is found to be

$$\det T = \Sigma t_\mu{}^4 \pm 2\Sigma t_\mu{}^2 t_\nu{}^2 + 8\Sigma t_{\mu\sigma}t_{\mu\tau}t_{\nu\sigma}t_{\nu\tau} \pm 8\Sigma t_{\mu\nu}t_{\sigma\tau}t_{\lambda\rho}t_{16}. \quad (3\cdot62)$$

Here the first two terms on the right are written in single-suffix notation,

† *Journ. Lond. Math. Soc.* **7**, 93, 272 (1932).

‡ Matrices of the form $E_\mu F_\nu$ are treated fully in Chapter x.

and the last two terms in double-suffix notation. In the second term the sign is positive if E_μ, E_ν anticommute and negative if they commute; in the last term the sign is positive if $\mu, \nu, \sigma, \tau, \lambda, \rho$ is an even permutation of the suffixes 0, 1, 2, 3, 4, 5 and negative for an odd permutation. It is understood that each component is written in one way only in the double-suffix notation; e.g. t_{21} may also be written as $-t_{12}$, but it is not to be included a second time in the summation on account of the two ways of denoting it.

The determinant of a mixed tensor is unaltered by tensor transformations. Thus (3·62) will be invariant when the special matrix frame used in (3·61) is changed to any other frame E_μ'.

We define the *determinant of an E-number* to be the function (3·62) of its sixteen coefficients. With this definition the determinant of an E-number is the same as the determinant of any matrix representation of it; and properties of fourfold matrices which involve their determinants can be extended to general symbolic E-numbers.

It is well known that the condition that a matrix T shall be singular is

$$\det T = 0. \tag{3·63}$$

Also, by a well-known theorem, for any two matrices S and T

$$\det (ST) = \det S \times \det T. \tag{3·64}$$

We see by inspection that when $E_\mu = S_a D_b$

$$\det E_\mu = \det S_a \times \det D_b = 1; \tag{3·65}$$

and, since the determinant is invariant for tensor transformations, this holds for the matrices E_μ of any complete orthogonal set. (The same result is also found directly from (3·62).)

By (3·64) and (3·65) $$\det (E_\mu T) = \det T. \tag{3·66}$$

And by (3·62) $$\det (a + bE_\mu) = (a^2 + b^2)^2 \quad (\mu \neq 16).$$

Hence

$$\det (\cos\theta + E_\mu \sin\theta) = 1 \quad (\mu \neq 16). \tag{3·671}$$

We generally write $\cos\theta + E_\mu \sin\theta = e^{E_\mu \theta}$ (see § 4·1); hence by (3·64)

$$\det (T e^{E_\mu \theta}) = \det T \quad (\mu \neq 16). \tag{3·672}$$

We shall later consider transformations of the form $T \to T e^{E_\mu \theta_\mu}$, $T \to e^{\frac{1}{2}E_\mu \theta_\mu} T e^{\frac{1}{2}E_\mu \theta_\mu}$, etc. By (3·672) any number of these transformations leaves $\det T$ invariant, if the algebraic transformation $\mu = 16$ is excluded. Further, if $\mu = 16$ and θ_{16} is real, the transformation does not alter the modulus $|\det T|$. A transformation which leaves $|\det T|$ unaltered is called a *unitary transformation*. Of the 32 possible transformations $e^{E_\mu \theta_\mu}$ (counting real and imaginary θ_μ as different transformations) the only one which is not unitary is that given by imaginary θ_{16}.

These results apply to general E-symbols as well as to matrices, the determinant being defined by (3·62).

3·7. Eigensymbols and Eigenvalues.

If X is any symbol, and ϕ is a symbol (not zero) such that

$$X\phi = \alpha\phi, \tag{3·71}$$

where α is an algebraic number, ϕ is said to be an *eigensymbol* of X and α an *eigenvalue* of X. We collect here some of the most important properties of eigensymbols. The results are given for final eigensymbols; but there are in all cases corresponding theorems for initial eigensymbols defined by $\phi X = \alpha\phi$.

If f is a polynomial function, repeated application of (3·71) gives

$$f(X).\phi = f(\alpha).\phi. \tag{3·715}$$

(*a*) If the symbol X satisfies a polynomial equation $f(X)=0$, the only possible eigenvalues of X are roots of $f(\alpha)=0$. For we have $f(X).\phi=0$, and therefore $f(\alpha).\phi=0$. Then, since ϕ is not zero, $f(\alpha)=0$.

In particular if X^2 is algebraic and equal to m^2, X has only two possible eigenvalues $\pm m$. The eigenvalues of the E_μ are $\pm i$.

(*b*) A symbol which has an eigenvalue 0 has no reciprocal. For if $X\phi=0$ and $X^{-1}X=1$, we have $0=X^{-1}(X\phi)=\phi$. But an eigensymbol, by definition, is not zero.

Hence if an E-number or matrix T has a zero eigenvalue, it is singular, and $\det T=0$.

(*c*) If T has an eigenvalue λ, $T-\lambda$ has an eigenvalue 0, so that

$$\det(T-\lambda)=0. \tag{3·72}$$

Accordingly the eigenvalues λ of an E-number or matrix are the roots of equation (3·72), which is called the *characteristic equation*. For an E-number or fourfold matrix the characteristic equation is of the fourth degree in λ, and may be written

$$f(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)(\lambda-\lambda_3)(\lambda-\lambda_4)=0. \tag{3·73}$$

It is known that a matrix satisfies its own characteristic equation (Hamilton-Cayley theorem), so that we have

$$f(T)=(T-\lambda_1)(T-\lambda_2)(T-\lambda_3)(T-\lambda_4)=0. \tag{3·74}$$

Equation (3·74) may be regarded as the converse of the result (*a*).

The polynomial equation of lowest degree satisfied by a symbol T is called the *minimum equation*. Since every eigenvalue must be a root of the minimum equation (by (*a*)) and every root of the characteristic equation is an eigenvalue, the minimum equation can only differ from the characteristic equation if the latter has repeated roots.

(*d*) To find an eigensymbol of T corresponding to one of its eigenvalues, say λ_1, we proceed as follows. Let $m(T)=0$ be the minimum equation, and let $g(T)=m(T)/(T-\lambda_1)$. Since $(T-\lambda_1)$ is a factor of $m(T)$, $g(T)$ is a poly-

nomial in T; it cannot vanish, because $g(T)=0$ would be a polynomial equation of degree lower than the minimum equation. Let $\phi=g(T).\chi$, where χ is any symbol. Then

$$(T-\lambda_1)\phi=(T-\lambda_1)g(T).\chi=m(T).\chi=0.$$

Hence $T\phi=\lambda_1\phi$; so that ϕ is the required eigensymbol.

(*e*) Mutually commuting matrices S, T, U, ... have a common eigensymbol, and in particular a common eigenvector.

We form polynomials $g_1(S)$, $g_2(T)$, $g_3(U)$ for S, T, U as in (*d*), and take $\phi=g_1(S)g_2(T)g_3(U)\chi$. Since $g_1(S)$, $g_2(T)$, $g_3(U)$ commute, it follows as in (*d*) that ϕ is an eigensymbol of S and T and U. Also $g_1(S)$, $g_2(T)$, $g_3(U)$ are matrices, and their product is a matrix; hence if χ is a vector, ϕ will be a vector.

(*f*) If X and Y have a common eigensymbol ϕ, we have $(XY-YX)\phi=0$, so that $XY-YX$ is either singular or zero.

(*g*) If E_μ and E_ν commute, and ϕ is an eigensymbol of $aE_\mu+bE_\nu$, where $a^2\neq b^2$, then ϕ is an eigensymbol of E_μ and E_ν.

For if
$$(aE_\mu+bE_\nu)\phi=\alpha\phi,$$
we have, on multiplying by $aE_\mu-bE_\nu$,
$$(b^2-a^2)\phi=\alpha(aE_\mu-bE_\nu)\phi,$$
so that ϕ is also an eigensymbol of $aE_\mu-bE_\nu$. Hence ϕ is an eigensymbol of $(aE_\mu+bE_\nu)\pm(aE_\mu-bE_\nu)$, i.e. of E_μ and E_ν.

(*h*) If E_μ, E_ν, E_σ mutually commute, and ϕ is an eigensymbol of
$$aE_\mu+bE_\nu+cE_\sigma,$$
where $a^2\neq b^2\neq c^2$, then ϕ is an eigensymbol of E_μ, E_ν and E_σ.

By (2·36) and (2·343) $E_\sigma=\pm iE_\mu E_\nu$. Hence, if $c'=\pm ic$, the datum is
$$(aE_\mu+bE_\nu+c'E_\mu E_\nu-\alpha)\phi=0.$$
We obtain, on multiplication by E_μ, E_ν, $E_\mu E_\nu$,
$$(-\alpha E_\mu-c'E_\nu+bE_\mu E_\nu-a)\phi=0,$$
$$(-c'E_\mu-\alpha E_\nu+aE_\mu E_\nu-b)\phi=0,$$
$$(-bE_\mu-aE_\nu-\alpha E_\mu E_\nu+c')\phi=0.$$
Hence the determinant
$$\begin{vmatrix} a & b & c' & -\alpha \\ -\alpha & -c' & b & -a \\ -c' & -\alpha & a & -b \\ -b & -a & -\alpha & c' \end{vmatrix}=0.$$
This gives the possible eigenvalues α. Then, by eliminating E_ν and $E_\mu E_\nu$ from three of the equations, we obtain a result of the form $(jE_\mu+k)\phi=0$. Hence if $j\neq 0$, ϕ is an eigensymbol of E_μ.

The singular case, when $j=0$ for every combination of three equations, occurs when the four minors of the fourth column of the above determinant all vanish. It is easily found that this condition requires that two of the quantities a^2, b^2, $-c'^2$ shall be equal. This is excluded by the enunciation

(*i*) If X, Y, Z are commuting symbols with eigenvalues x_i, y_i, z_i $(i=1, 2, 3, \ldots)$, the eigenvalues of any rational function $f(X, Y, Z)$ are included among the quantities $f(x_i, y_j, z_k)$.†

We can write

$$\begin{aligned} f(X,Y,Z)-f(x_i,y_j,z_k) &= \{f(X,Y,Z)-f(x_i,Y,Z)\} \\ &\quad +\{f(x_i,Y,Z)-f(x_i,y_j,Z)\}+\{f(x_i,y_j,Z)-f(x_i,y_j,z_k)\} \\ &= A_{ijk}(X-x_i)+B_{ijk}(Y-y_j)+C_{ijk}(Z-z_k), \end{aligned} \tag{3·75}$$

since the first bracket vanishes when $X=x_i$, the second when $Y=y_j$, the third when $Z=z_k$. Now form the product

$$\prod_{i,j,k}\{f(X,Y,Z)-f(x_i,y_j,z_k)\}$$

for all combinations of values of i, j, k. By substituting (3·75) in it, we express it as the sum of a number of terms containing products

$$(X-x_1)^r(X-x_2)^s\ldots(Y-y_1)^t(Y-y_2)^u\ldots(Z-z_1)^v(Z-z_2)^w\ldots.$$

Every term will contain a complete set of eigenvalues of at least one of the symbols X, Y, Z. For, if not, let there be a term which does not contain the factors $(X-x_l)$, $(Y-y_m)$, $(Z-z_n)$. But one of the factors is

$$A_{lmn}(X-x_l)+B_{lmn}(Y-y_m)+C_{lmn}(Z-z_n),$$

so that either $(X-x_l)$ or $(Y-y_m)$ or $(Z-z_n)$ must appear in every term.

The minimum equation for X is $m(X)=\Pi_i(X-x_i)=0$. Since every term contains $m(X)$ or $m(Y)$ or $m(Z)$, every term vanishes; and we have

$$\prod_{i,j,k}\{f(X,Y,Z)-f(x_i,y_j,z_k)\}=0. \tag{3·76}$$

This is a polynomial equation satisfied by the symbol $f(X, Y, Z)$, and its roots $f(x_i, y_j, z_k)$ accordingly include all possible eigenvalues of the symbol. Not every root will be an eigenvalue; for example, if X, Y, Z are fourfold matrices, $f(X, Y, Z)$ will be a fourfold matrix, so that not more than 4 of the 64 roots of (3·76) can be eigenvalues.

Of the above results (*a*), (*b*), (*d*), (*e*), (*f*), (*i*) apply to all symbols which satisfy a polynomial equation. A common example of a symbol which does not satisfy any polynomial equation is $\partial/\partial x$.

† Frobenius's theorem. It holds for any number of commuting symbols; we here take three as a sufficient illustration.

3·8. Pauli Matrices.

A notation for the E-symbols based upon the conjugate triads (2·37) is sometimes useful.† Denote the symbols $E_{\mu\nu}, E_{\nu\sigma}, E_{\sigma\mu}; E_{\tau\lambda}, E_{\lambda\rho}, E_{\rho\tau}$, which form conjugate triads, by

$$A_1, A_2, A_3; \quad B_1, B_2, B_3.$$

Then the sixteen E_μ can be written as

$$A_\sigma, B_\sigma, iA_\sigma B_\tau, i \quad (\sigma, \tau = 1, 2, 3). \tag{3·81}$$

Each A is the product of the other two A's, and each B is the product of the other two B's. The rules of commutation are: an A anticommutes with an A, and a B with a B; an A commutes with a B.

There is one pair of conjugate triads in which all six matrices are real, viz.

$$E_{23}, E_{31}, E_{12}; \quad E_{45}, -E_5, E_4, \tag{3·82}$$

E_4, E_5 being the real matrices of the pentad. The other ten matrices of the set are imaginary. With this identification, the real or imaginary character of the matrices in (3·81) is explicitly indicated by the absence or presence of i.

This method of constructing a complete set can be exhibited in another way. We apply the treatment of § 2·2 to two symbols instead of four. Let A_1, A_2 be any two symbols which satisfy

$$A_\mu{}^2 = -1, \quad A_\mu A_\nu = -A_\nu A_\mu. \tag{3·83}$$

Then, if $A_3 = A_1 A_2$, A_3 is an additional symbol satisfying (3·83). The symbols

$$A_1, A_2, A_3, i \tag{3·84}$$

form a "minor complete set". Calling any linear function of them an A-number, the operations of addition, subtraction and multiplication applied to A-numbers always yield A-numbers.

A minor complete set can be represented by twofold matrices. Let

$$\zeta_1 = \begin{matrix} i & 0 \\ 0 & -i \end{matrix}, \qquad \zeta_2 = \begin{matrix} 0 & i \\ i & 0 \end{matrix}, \qquad \zeta_3 = \begin{matrix} 0 & -1 \\ 1 & 0 \end{matrix}. \tag{3·85}$$

These satisfy

$$\zeta_\mu{}^2 = -1, \quad \zeta_\mu \zeta_\nu = \zeta_\lambda = -\zeta_\nu \zeta_\mu, \tag{3·86}$$

where μ, ν, λ are in cyclic order. Thus $\zeta_1, \zeta_2, \zeta_3, i$ is a particular representation of A_1, A_2, A_3, i. The matrices (3·85) are called *Pauli matrices*.‡

We can show, in the same way as for the E-symbols, that with this identification every A-number is represented by a twofold matrix and every twofold matrix can be expressed as an A-number. We find also a transformation theory for minor complete sets analogous to § 2·7, viz.

$$\zeta_\mu' = P \zeta_\mu P^{-1},$$

where

$$P = \alpha(-1 + \zeta_1'\zeta_1 + \zeta_2'\zeta_2 + \zeta_3'\zeta_3).$$

† This was pointed out by G. Lemaître.

‡ Most writers employ the matrices $i\zeta_\mu$ whose squares are $+1$.

In Case (*b*), in which the ζ_μ' are new symbols commuting with the ζ_μ, we have $\alpha = \frac{1}{2}$, so that the transformation operator is

$$P = P^{-1} = \tfrac{1}{2}(-1 + \zeta_1\zeta_1' + \zeta_2\zeta_2' + \zeta_3\zeta_3'). \qquad (3{\cdot}87)$$

To reproduce the sixteen E_μ we require two minor complete sets A and B. These may be represented by two sets of Pauli matrices ζ_μ, θ_μ, provided that the products $\zeta_\mu\theta_\nu$ are taken as outer products. The outer product of two twofold matrices is a fourfold matrix. We are thus led back to the ordinary representation of the E-symbols by fourfold matrices.

3·9. Left-handed Frames.

The coefficients t_μ of an E-number are in general complex algebraic numbers, say $\tau_\mu + i\sigma_\mu$. For i we have a choice of two algebraic square roots of -1, which we shall call i_1, i_2. We may regard the symbols $(1, i)$ as constituting an algebraic frame; we have then to distinguish two possible algebraic frames $(1, i_1)$, $(1, i_2)$ either of which can be combined with a given symbolic frame E_μ.

We have hitherto used E_{16} and i indiscriminately; but it is now desirable to define the structure of an orthogonal symbolic frame unambiguously by eliminating i in the fundamental equations. We therefore replace (2·22) and (2·343) by

$$E_{16}E_5 = E_1E_2E_3E_4, \quad E_{\mu\nu}E_{\sigma\tau} = E_{16}E_{\lambda\rho}, \qquad (3{\cdot}91)$$

where $\mu, \nu, \sigma, \tau, \lambda, \rho$ is an even permutation of 0, 1, 2, 3, 4, 5.

An E-number $T = \Sigma(\tau_\mu + i\sigma_\mu)E_\mu$ will involve an algebraic square root of -1 denoted by E_{16} which occurs in the symbolic frame, and also an algebraic square root of -1 denoted by i which occurs in the coefficients. The complete reference frame for E-numbers thus consists of

(*a*) A symbolic frame E_μ,

(*b*) An algebraic frame $(1, i)$ for the coefficients.

E_{16} and i may or may not be the same root of -1. *Absolutely* it is meaningless to inquire whether they are the same root. But if we (arbitrarily) regard them as the same in one complete reference frame, we can define other complete reference frames in which they are opposite roots.

A complete reference frame in which $E_{16} = i$ will be called *right-handed*, and a frame in which $E_{16} = -i$ will be called *left-handed*.

Since E_{16} is invariant for tensor transformations, there is no transformation of the type $F_\mu = qE_\mu q'$ between right- and left-handed frames.† Their relation is like that of right- and left-handed systems of rectangular coordinates, which cannot be changed into one another by rotation.

We usually treat the algebraic frame as unalterable so that a right-handed frame E_μ and a left-handed frame F_μ are distinguished by $F_{16} = -E_{16}$.

† The existence of two kinds of complete sets not transformable into one another by relativity transformation was, I think, first recognised by S. R. Milner.

The relation of the other matrices can have various forms; we give the three most important. If the frames are constructed from the same tetrad

we have by (3·91) $$E_1, E_2, E_3, E_4 = F_1, F_2, F_3, F_4,$$

$$F_{16}, F_5, F_{15}, F_{25}, F_{35}, F_{45} = -E_{16}, -E_5, -E_{15}, -E_{25}, -E_{35}, -E_{45}, \tag{3·92}$$

the other ten matrices being the same. Similarly if they have the tetrad $E_{15}, E_{25}, E_{35}, E_{45}$ in common,

$$F_{16}, F_1, F_2, F_3, F_4, F_5 = -E_{16}, -E_1, -E_2, -E_3, -E_4, -E_5, \tag{3·93}$$

the other ten matrices being the same. Another form, obtained by giving alternative signs to i in (3·81) is

$$F_r = E_r, \quad F_i = -E_i, \tag{3·94}$$

where E_r denotes the real matrices $E_{23}, E_{31}, E_{12}, E_4, E_5, E_{45}$, and E_i the imaginary matrices. In this last case we change from a right- to a left-handed set by writing $-i$ for i in the elements of the matrices.

We shall find later that the distinction between positive and negative electric charges corresponds to the distinction between right- and left-handed frames; so that a positive charge cannot be changed into a negative charge by a relativity transformation of frame.

CHAPTER IV

SPACE VECTORS

4·1. Rotations.

When an exponential contains non-algebraic symbols, it is understood to be defined by the exponential series. Thus

$$e^{E_\mu\theta} = 1 + E_\mu\theta + \tfrac{1}{2}E_\mu{}^2\theta^2 + \tfrac{1}{6}E_\mu{}^3\theta^3 + \tfrac{1}{24}E_\mu{}^4\theta^4 + \ldots$$
$$= (1 - \tfrac{1}{2}\theta^2 + \tfrac{1}{24}\theta^4 - \ldots) + E_\mu(\theta - \tfrac{1}{6}\theta^3 + \ldots),$$

since $E_\mu{}^2 = -1$. Hence

$$e^{E_\mu\theta} = \cos\theta + E_\mu \sin\theta. \tag{4·11}$$

In fact, so long as no opportunity for exhibiting non-commutative properties arises, E_μ is indistinguishable from i. The reciprocal of $e^{E_\mu\theta}$ is $e^{-E_\mu\theta}$.

The ordinary factorisation of an exponential $e^{\alpha+\beta} = e^\alpha . e^\beta$ holds only so long as α and β commute. For example,

$$e^{E_1\theta + E_{23}\phi} = e^{E_1\theta} . e^{E_{23}\phi}, \quad e^{E_1\theta + E_2\phi} \neq e^{E_1\theta} . e^{E_2\phi}. \tag{4·115}$$

By (4·11)

$$e^{E_\mu\theta}E_\nu = (\cos\theta + E_\mu\sin\theta)\,E_\nu$$
$$= E_\nu(\cos\theta + E_\mu\sin\theta) \quad \text{if } E_\mu,\ E_\nu \text{ commute}$$
$$= E_\nu(\cos\theta - E_\mu\sin\theta) \quad \text{if } E_\mu,\ E_\nu \text{ anticommute.}$$

Hence

$$\left.\begin{aligned} e^{E_\mu\theta}E_\nu &= E_\nu e^{E_\mu\theta} \quad \text{if } E_\mu,\ E_\nu \text{ commute} \\ &= E_\nu e^{-E_\mu\theta} \quad \text{if } E_\mu,\ E_\nu \text{ anticommute} \end{aligned}\right\}. \tag{4·12}$$

We now consider the relativity transformations of the E-numbers describing a physical system. By (2·94) the change $t_\mu \to t_\mu'$ of the coefficients, due to a rotation of the physical system relative to a fixed frame E_μ, is given by

$$\Sigma t_\mu' E_\mu = q\,(\Sigma t_\mu E_\mu)\,q'.$$

Let

$$q = e^{\frac{1}{2}E_{12}\theta}, \quad q' = e^{-\frac{1}{2}E_{12}\theta},$$

which satisfies $qq' = 1$. Then

$$\Sigma t_\mu' E_\mu = e^{\frac{1}{2}E_{12}\theta}\,(\Sigma t_\mu E_\mu)\,e^{-\frac{1}{2}E_{12}\theta} \tag{4·13}$$
$$= \Sigma_\alpha t_\mu E_\mu + \Sigma_\beta t_\mu E_\mu e^{-E_{12}\theta}, \tag{4·14}$$

by (4·12), where Σ_α denotes summation of the eight terms which commute with E_{12}, and Σ_β summation of the eight terms which anticommute. As an example of terms anticommuting with E_{12} and therefore included in Σ_β, we take $t_1E_1 + t_2E_2$. We have

$$(t_1E_1 + t_2E_2)\,e^{-E_{12}\theta} = (t_1E_1 + t_2E_2)\cos\theta - (t_1E_1 + t_2E_2)\,E_{12}\sin\theta$$
$$= (t_1E_1 + t_2E_2)\cos\theta - (-t_1E_2 + t_2E_1)\sin\theta$$
$$= (t_1\cos\theta - t_2\sin\theta)\,E_1 + (t_1\sin\theta + t_2\cos\theta)\,E_2.$$

By § 2·4 (*b*) we may equate coefficients of the same matrix on both sides of (4·14). Hence

$$t_1' = t_1 \cos\theta - t_2 \sin\theta, \quad t_2' = t_1 \sin\theta + t_2 \cos\theta. \tag{4·15}$$

That is to say, the relativity transformation $q = e^{\frac{1}{2}E_{12}\theta}$ rotates (t_1, t_2) through an angle θ.

Examining the other terms in Σ_β we find that three other pairs of components are rotated in the same way, viz. t_{13}, t_{23}; t_{14}, t_{24}; t_{15}, t_{25}. The remaining components are unchanged.

Considering the more general relativity transformation $q = e^{\frac{1}{2}E_\mu\theta}$, and taking $\mu = 1, 2, \ldots 15$, we obtain 15 independent rotations of this type, each rotating four pairs of components. We may, if we like, add the E_{16} rotation, viz. $q = e^{\frac{1}{2}i\theta}$, $q' = e^{-\frac{1}{2}i\theta}$, but this does not alter T.

Two terms which mutually rotate necessarily anticommute with each other as well as with the transformation matrix. We can always find a transformation matrix which will rotate two anticommuting terms $E_{\mu\nu}$, $E_{\mu\sigma}$, viz. their product $E_{\nu\sigma}$. There is no corresponding mutual rotation of two commuting terms. Suppose, for example, that we try to rotate $E_1 t_1$ and $E_{23} t_{23}$. Since $E_{23} = -iE_{45}E_1$, we shall require the transformation $q = e^{\frac{1}{2}E_{45}\theta}$; but E_1 and E_{23} commute with E_{45}, and therefore come under Σ_α. Accordingly t_1, t_{23} are unchanged by the transformation.

Pairs of components which can rotate with one another will be called *perpendicular*; pairs which cannot rotate will be called *antiperpendicular*. We have the rule:

Matrices commute: components antiperpendicular;

Matrices anticommute: components perpendicular.

The term *orthogonal* will be understood to include antiperpendicularity as well as perpendicularity.

If the components t_μ are represented as coordinates in a space of 16 dimensions, the space contains certain planes in which rotation is possible and certain planes in which it is forbidden. Actually the 120 coordinate planes consist of 60 planes of rotation and 60 forbidden planes. More precisely we should say that rotation in a forbidden plane is a non-relativistic change; although depicted graphically as a rotation, it is an intrinsic deformation of the physical system described by t_μ.

Rotation in a forbidden plane will be called an *antiperpendicular rotation*. It is produced by the (non-relativistic) transformation

$$T' = qTq, \tag{4·16}$$

or

$$\Sigma t_\mu' E_\mu = q(\Sigma t_\mu E_\mu) q.$$

Taking $q = e^{\frac{1}{2}E_{12}\theta}$ as before, we have instead of (4·14)

$$\Sigma t_\mu' E_\mu = \Sigma_\alpha t_\mu E_\mu e^{E_{12}\theta} + \Sigma_\beta t_\mu E_\mu.$$

Hence pairs of terms which are not rotated by the relativity rotation are rotated by the antiperpendicular rotation, and *vice versa*. As an example of terms commuting with E_{12} and therefore included in Σ_α, we take

$$\begin{aligned}(t_3E_3+t_{45}E_{45})\,e^{E_{12}\theta} &= (t_3E_3+t_{45}E_{45})\cos\theta+(t_3E_3+t_{45}E_{45})\,E_{12}\sin\theta\\ &= (t_3E_3+t_{45}E_{45})\cos\theta+(it_3E_{45}+it_{45}E_3)\sin\theta\\ &= (t_3\cos\theta+it_{45}\sin\theta)\,E_3-i\,(-t_3\sin\theta+it_{45}\cos\theta)\,E_{45}.\end{aligned}$$

Hence $$t_3'=t_3\cos\theta+it_{45}\sin\theta,\quad it_{45}'=-t_3\sin\theta+it_{45}\cos\theta. \tag{4·17}$$

The other pairs rotated by the same transformation are

$$t_4,\ it_{35};\quad t_5,\ it_{34};\quad t_{12},\ it_{16}.$$

If alternatively we treat (4·16) as a transformation of the *frame*, so that $T'=\Sigma t_\mu E_\mu'$, where

$$E_\mu'=qE_\mu q, \tag{4·18}$$

the new frame E_μ' does not satisfy the conditions (2·23) for a complete orthogonal set. Its structure is intrinsically different from that of the standard frame E_μ; and the physical structure built by t_μ in such a frame is therefore not equivalent to the structure built by the same t_μ in an orthogonal frame. It is to be noticed that this argument that antiperpendicular rotations are non-relativistic changes, does not introduce the question whether (4·16) is a tensor transformation. We shall see later (Chapter VII) that in some cases (4·16) is a rather simple tensor transformation, though it is not that of a mixed tensor. I think that confusion of thought has often been caused by failure to recognise that, although tensor calculus is an almost indispensable tool in relativity theory, it does not in itself imply any relativistic hypothesis.

4·2. Alternative Treatment.

The rotations of the components t_μ can also be found directly from (3·35) or (3·37). Consider first a factorisable matrix $J=\psi\chi^*=\Sigma j_\mu E_\mu$. By (1·461) and (1·462) the relativity transformation $q=e^{\frac{1}{2}E_\mu\theta_\mu}$ gives

$$\psi'=e^{\frac{1}{2}E_\mu\theta_\mu}\psi,\quad \chi^{*\prime}=\chi^*e^{-\frac{1}{2}E_\mu\theta_\mu}.$$

Hence, by (3·37),

$$j_\nu'=-\tfrac{1}{4}\chi^{*\prime}E_\nu\psi'=-\tfrac{1}{4}\chi^*e^{-\frac{1}{2}E_\mu\theta_\mu}E_\nu e^{\frac{1}{2}E_\mu\theta_\mu}\psi.$$

If E_μ, E_ν anticommute, this becomes, by (4·12),

$$j_\nu'=-\tfrac{1}{4}\chi^*E_\nu e^{E_\mu\theta_\mu}\psi=-\tfrac{1}{4}\chi^*(E_\nu\cos\theta+E_\sigma\sin\theta)\psi,$$

where $E_\sigma=E_\nu E_\mu$. Hence, by (3·37),

$$j_\nu'=j_\nu\cos\theta+j_\sigma\sin\theta. \tag{4·21}$$

If E_μ, E_ν commute, the result is $j_\nu'=j_\nu$. The components of a general matrix T are transformed according to the same formula, because the general matrix can be expressed as the sum of a number of factorisable matrices.

It is tempting to combine these elementary rotations into a "general rotation" $q = e^{\frac{1}{2}\Sigma E_\mu \theta_\mu}$; but we have seen in (4·115) that non-commuting exponentials cannot be compounded in this way. This prohibition merely reflects the non-commutability of rotations in ordinary Euclidean geometry, and is not a peculiarity of the t-space. The objection, however, does not apply to a combination of infinitesimal rotations. The most general infinitesimal rotation, corresponding to a general infinitesimal matrix $d\Theta$, is

$$q = e^{\frac{1}{2}d\Theta} = e^{\frac{1}{2}\Sigma E_\mu d\theta_\mu}, \tag{4·22}$$

the squares and products of $d\theta_\mu$ being neglected, so that the question of non-commutation does not arise. Equivalently (4·22) may be written

$$q = 1 + \tfrac{1}{2}d\Theta = 1 + \tfrac{1}{2}\Sigma E_\mu d\theta_\mu. \tag{4·23}$$

When the t_μ are represented as coordinates in a 16-dimensional space, each elementary relativity rotation $q = e^{\frac{1}{2}E_\mu \theta_\mu}$ appears as a rotation through an angle θ_μ occurring simultaneously in four different planes. The 15 values of μ (excluding $\mu = 16$) accordingly give 60 planes of rotation; and there remain 60 forbidden planes. It is evident that the geometry of this 16-space is different from that of any type of space ordinarily studied. The novel features are (1) the occurrence of antiperpendicular pairs of axes, and (2) the locking together of four rotations. We shall now consider how to represent the components t_μ in a more familiar type of space.

4·3. Five-dimensional Euclidean Space.

A pentad provides five mutually perpendicular components t_1, t_2, t_3, t_4, t_5. Any pair of these can be rotated. Moreover, the rotation matrix E_{12} commutes with E_3, E_4, E_5, so that the mutual rotation of t_1 and t_2 leaves t_3, t_4, t_5 unaltered. It is therefore a simple Euclidean rotation so far as these five components are concerned.

If then we represent $(t_1, t_2, t_3, t_4, t_5)$ as coordinates (or as components of a vector) in five-dimensional space, this space has the ordinary relativistic properties of Euclidean space, namely that simple rotation in any of the ten coordinate planes is a relativistic transformation. Accordingly the intrinsic properties of a physical system are not affected by changing its orientation in this space.

Thus by limiting ourselves to a sub-space of five dimensions we encounter neither of the complications of the geometry of 16-space. The domain of $(t_1, t_2, t_3, t_4, t_5)$ has the properties which we attribute to ordinary space; and (leaving aside the fifth dimension for the present) we may identify physical space-time with a continuum constructed in this way.

This provides a linkage between wave tensors and space tensors (§ 1·2). *A space vector is the pentadic part of a mixed wave tensor.* Ordinarily we regard the components of a space vector as an array $(t_1, t_2, t_3, t_4, t_5)$, whereas in a

wave tensor they are strung together with symbolic coefficients as a linear expression $t_1E_1+t_2E_2+t_3E_3+t_4E_4+t_5E_5$; but the latter mode of representing space vectors has long been recognised as permissible, e.g. in quaternion notation, so that the difference of form need not be stressed.

This is not a hypothetical identification. In § 1·2 we left the basic wave vector undefined, and we are therefore free to define it at this stage. We now *define* the relation of wave tensors to the ordinary space vectors of physics to be such as is expressed by this identification. Henceforth our calculus embraces both wave tensors and space tensors.

The question remains, What is the significance of the eleven remaining components of the wave tensor? When the 5-vector undergoes a relativity rotation, e.g. in the plane t_1t_2, rotations also occur between t_{13}, t_{23}; t_{14}, t_{24}; t_{15}, t_{25}. The ordinary tensor calculus provides for these locked transformations; a transformation consequent on a transformation of the basal space vector is made automatic by assigning the appropriate space tensor character to the quantities concerned in it. We shall prove in § 4·5 that the following is the required specification:

(*a*) t_1, t_2, t_3, t_4, t_5 is a space vector,

(*b*) $t_{12}, t_{13}, \ldots t_{45}$ is a 10-vector or antisymmetrical space tensor of the second rank (analogous to a 6-vector in four dimensions),

(*c*) t_{16} is an invariant.

The statement that (*a*) is a vector and (*b*) an antisymmetrical tensor of the second rank secures that the transformations of (*b*) are locked to those of (*a*) in such a way that the rotations of pairs of terms occur in groups of four as required.

We shall call the group of space tensors (*a*), (*b*), (*c*) a *complete space vector*, or (if no ambiguity is likely to arise) simply a *space vector*. We have therefore the simple relation

$$\text{Mixed wave tensor} = \text{Complete space vector.} \qquad (4\cdot31)$$

In dealing with space vectors we recognise only ten relativity rotations, viz. the ten rotations of five-dimensional space. The wave tensor had 15 relativity rotations (or 16, if we count the algebraic rotation). The five extra rotations intermingle the two space tensors (*a*) and (*b*); for example, the E_2 rotation rotates the component t_1 of (*a*) with the component t_{12} of (*b*). Ordinarily (*a*) and (*b*) will be recognised in physics under different names, e.g. velocity and spin, and a transformation which does not keep them distinct could only be pictured at the cost of abandoning the usual representation in space and time. We rather miss the point of the interpretation of mixed wave tensors as space vectors, if we go on to attribute to the space vectors transformation properties outside those which the name ordinarily suggests. There is no advantage in introducing space vectors so generalised that we

can no longer "see" them in space; for if we are not going to use the space picture, we have to fall back on their analytical description as mixed wave tensors, and it is more appropriate to refer to them by that name. We may therefore agree that the name space vector implies that only the ordinary relativity rotations of the space are under consideration; and that if it is desired to include the transformations which transcend space-time representation the proper designation is wave tensor.

This question of nomenclature arises because in § 7·6 we shall define an important entity which behaves as a mixed wave tensor for the ten rotations of 5-space, but not for the other rotations. Thus it may be properly described as a space vector, but not as a mixed wave tensor. Except in this connection it is not necessary to emphasise the distinction, and I shall generally use the two names as equivalent.

4·4. Four-dimensional Spherical Space.

We have now to consider why the actual world is four-dimensional, although the analytical theory (which we have reason to think is appropriate to the physical world) seems to provide for five dimensions. The answer is not difficult to find. Space-time is four-dimensional, but it is not flat (Euclidean); and if we restrict ourselves to Euclidean geometry, we require at least one more dimension to represent its curvature.

Our actual space-time with its irregular curvature, due to local gravitational fields, requires a ten-dimensional Euclidean space for its representation; but we do not arrive at this complexity until we provide for a more varied content of the universe than can be represented by a simple wave tensor. Any problem which is mathematically simple is necessarily highly idealised; and the simple wave tensor with which we commence our study can carry us only a little way towards actuality. We have found that it is capable of rotation, without intrinsic change, in all planes in five dimensions. This leads immediately to the conception of an entity distributed over a continuum which is a hypersphere (four-dimensional manifold) in a Euclidean space of five dimensions. We must start with this simplified space-time, and watch the more complex characteristics of actual space-time grow out of it as the theory develops.

We shall first show that it is the hypersphere (and not the five-dimensional space in which it is represented) which constitutes the *physical* continuum. This is because the hypersphere is a locus of equivalent points, whereas the points in five-dimensions are not generally "equivalent" points.

Let the coordinates of a point P be $(t_1, t_2, t_3, t_4, t_5)$. Any of the ten rotations in five dimensions will (if it displaces P) carry P to a new position P' on the same hypersphere about the origin. Then the points P, P' are con-

structed according to the same specification in different but equivalent frames, and are therefore equivalent points (§ 2·9). None of the relativity transformations provides any connection between points which are not on the same hypersphere. If we consider a point P'' such that PP'' is normal to the hypersphere, P and P'' are not equivalent points. The transformation $(0, 0, 0, 0, t_5) \rightarrow (0, 0, 0, 0, t_5'')$ is not a relativity transformation; it is in fact an antiperpendicular rotation with matrix E_{16}.

Thus displacement normal to the hypersphere involves, either a different construction, or the same construction in a frame which differs intrinsically from the original frame. Taking the latter view, the difference between the two frames is evidently a difference of scale-constant. Our five-dimensional picture accordingly represents change of scale-constant or *gauge* by displacement in a fifth coordinate, normal to the four ordinary coordinates used to represent position. A similar graphical representation of change of gauge is used in "Projective Relativity".

It is the essence of the elementary conception of space and time that all points of it are equivalent. A particle is not intrinsically different because it is at a different point of space or because it is contemplated at a different time. It is true that in later developments regions of space and time are distinguished from one another by varying curvature; so that the space-time background is no longer a "blank sheet". But we have to trace the origin of these distinctions, and must begin with the blank sheet. This is provided by the hyperspherical continuum of equivalent points; the curvature is uniform, and every part of the continuum is precisely similar to every other part, as the conception of "equivalence" implies.

Thus although the transformation theory introduces the conception of five dimensions, it is clear from the start that there is an absolute distinction between the displacements lying in the four-dimensional hypersphere and displacements in the fifth dimension normal to the hypersphere. Whereas the former changes of a system are of the type which we conceive as displacement in physical space and time, the latter are only "displacements" in the sense in which we regard any change of quality of a system (scale, temperature, entropy, etc.) as a displacement of the representative point in a graph of that quality.

Let us now confine attention to a region of the hypersphere small enough to be treated as flat, and let t_5 be the coordinate normal to this region, the coordinates in space-time being (t_1, t_2, t_3, t_4). We now exclude the rotations $E_{15}, E_{25}, E_{35}, E_{45}$, because they would carry us away from the small region considered and mix the scale-coordinate t_5 with the coordinates recognised in our ordinary outlook. There remain the relativity transformations corresponding to the matrices

$$E_{23}, E_{31}, E_{12}, E_{14}, E_{24}, E_{34}. \qquad (4·41)$$

These constitute the transformations admitted in special relativity theory, viz. three rotations in space and three Lorentz transformations. We now consider how the general mixed wave tensor will appear from this outlook —the ordinary outlook on which current nomenclature is chiefly based.

The mixed wave tensor is found to break up into

$$\left.\begin{array}{ll}(a) & \text{a vector } t_1, t_2, t_3, t_4,\\ (b) & \text{another (adjoint) vector } t_{15}, t_{25}, t_{35}, t_{45},\\ (c) & \text{a 6-vector } t_{23}, t_{31}, t_{12}, t_{14}, t_{24}, t_{34},\\ (d) & \text{two invariants } t_5, t_{16}.\end{array}\right\} \qquad (4\cdot42)$$

The proof that (a) and (b) undergo the same transformations for any of the rotations (4·41), and are therefore "tensors of the same kind", is obvious. The proof that (c) is a 6-vector is given in § 4·5.

Thus in four dimensions the complete space vector is composed of two vectors, a 6-vector, and two invariants. These remain distinct in any of the six internal rotations of the 4-space—which transform the 4-space into itself. Naturally the use of the four-dimensional picture presupposes that we are not interested in the transformations which cannot be represented in the 4-space; if we have occasion to refer to them we must revert to one of the other modes of description.

If we are right in our belief that all physical phenomena are analysable into ultimate elements described by wave vectors and their combinations, it follows that an ordinary space vector cannot occur alone; it is part of a group of allied space tensors (4·42). The other members of the group may, of course, have zero value, but that is not the same as being non-existent. This is one of the ways in which the new outlook enriches the earlier theory.

4·5. Proof that $(t_{23}, t_{31}, t_{12}, t_{14}, t_{24}, t_{34})$ is a 6-vector.

If $x_\mu, y_\mu (\mu = 1, 2, 3, 4)$ are ordinary space vectors X, Y, their vector product consists of six quantities $(x_\mu y_\nu - x_\nu y_\mu)$. Then any set of six quantities which transform according to the same law as $(x_\mu y_\nu - x_\nu y_\mu)$ is called a 6-vector.

In matrix form the vectors are

$$X = E_1 x_1 + E_2 x_2 + E_3 x_3 + E_4 x_4,$$
$$Y = E_1 y_1 + E_2 y_2 + E_3 y_3 + E_4 y_4.$$

Their matrix product XY is found by direct multiplication (using $E_\mu{}^2 = -1$, $E_\mu E_\nu = -E_\nu E_\mu$) to be

$$XY = -(xy) + S, \qquad (4\cdot51)$$

where (xy) is the scalar product, and

$$S = E_{23}(x_2 y_3 - x_3 y_2) + E_{31}(x_3 y_1 - x_1 y_3) + \ldots + E_{34}(x_3 y_4 - x_4 y_3). \qquad (4\cdot52)$$

Apply a transformation q representing a rotation or Lorentz transformation in four dimensions. Then (xy) is invariant, and

$$X'Y' = qXq'qYq' = qXYq'.$$

Hence by (4·51)
$$S' = X'Y' + (xy) = qXYq' + (xy)$$
$$= q\{XY + (xy)\}q' = qSq'. \tag{4·53}$$

Let
$$U = E_{23}t_{23} + E_{31}t_{31} + \ldots + E_{34}t_{34}. \tag{4·54}$$

Then, since the constituents (a), (b), (c), (d) of the complete space vector T transform separately, we have
$$U' = qUq'. \tag{4·55}$$

By (4·53) and (4·55) U and S obey the same transformation law for rotations and Lorentz transformations in four dimensions. It follows that their components $t_{\mu\nu}$ and $(x_\mu y_\nu - x_\nu y_\mu)$ obey the same transformation law. Hence $t_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ is a 6-vector.

We can show similarly that $t_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4, 5)$ is a 10-vector in 5-space.

It will be seen that the matrix product XY, used in the present calculus, is not the same as the ordinary vector product $X \times Y$, but is $X \times Y - (xy)$. When two vectors are perpendicular their scalar product (xy) vanishes, and the matrix product is then the vector product.

The product of two complete space vectors is a complete space vector. For the product $W = UV$ transforms according to the law
$$W' = U'V' = qUq'qVq' = qUVq' = qWq'. \tag{4·56}$$

But, as we have seen above, the matrix product here employed is a type of combination which has no exact counterpart in the ordinary theory of vectors.

4·6. Volume Elements.

The expressions $E_1 dx_1$, $E_2 dx_2$, $E_3 dx_3$, $E_4 dx_4$ are the wave tensor notation for four space vectors (displacements) along the four rectangular axes in space-time. Their product constitutes a volume element dV_{1234}. Writing
$$d\tau = dx_1 dx_2 dx_3 dx_4, \qquad dV_{1234} = E_1 E_2 E_3 E_4 d\tau = iE_5 d\tau, \tag{4·61}$$
by (2·22). Since the E's anticommute, $dV_{\mu\nu\sigma\tau}$ is antisymmetrical in its four suffixes as in the ordinary tensor calculus.† The factor iE_5 corresponds to $\sqrt{-g}$. We are restricted to rectangular coordinates (one of them time-like), and therefore $-g$ is always 1; but we deviate from the usual theory which assumes that the radical indicates the algebraic square root. We take instead a matrix square root iE_5.

Consequently, notwithstanding that we use rectangular coordinates, the distinction between vectors and vector densities does not wholly disappear. In the ordinary calculus we have, corresponding to a vector A_μ, a vector density $A_\mu\sqrt{-g}$. Here $\sqrt{-g} = iE_5$, so that corresponding to a vector
$$(E_1 t_1 + E_2 t_2 + E_3 t_3 + E_4 t_4)$$
we have a vector density
$$i(E_{15}t_1 + E_{25}t_2 + E_{35}t_3 + E_{45}t_4).$$

† *Mathematical Theory of Relativity*, § 49.

Of the two adjoint vectors in (4·42) one represents a vector and the other a vector density. We shall show later (§ 5·8) how it is possible to decide which is which.

More generally we define the volume element contained by four space vectors d^1x_μ, d^2x_μ, d^3x_μ, d^4x_μ, which do not coincide with the axes, to be the *permanent*

$$dV_{1234}=\begin{bmatrix} E_1 d^1x_1, & E_2 d^1x_2, & E_3 d^1x_3, & E_4 d^1x_4 \\ E_1 d^2x_1, & E_2 d^2x_2, & E_3 d^2x_3, & E_4 d^2x_4 \\ E_1 d^3x_1, & E_2 d^3x_2, & E_3 d^3x_3, & E_4 d^3x_4 \\ E_1 d^4x_1, & E_2 d^4x_2, & E_3 d^4x_3, & E_4 d^4x_4 \end{bmatrix}. \tag{4·62}$$

A permanent is expanded like a determinant except that all the terms are given positive sign. The factors are arranged in the order of the rows from which they are taken; since the rows correspond to the four vectors, this is the natural order of the factors when the vectors are multiplied in a given order. The anticommutation of the four E's provides an alternation of sign which converts the permanent into a determinant; and (4·62) reduces to

$$dV_{1234}=E_1E_2E_3E_4\det(d^\alpha x_\beta)=iE_5\det(d^\alpha x_\beta).$$

In our later developments three-dimensional vector densities are more important than the foregoing four-dimensional densities. The volume-element of three-dimensional space, contained by vectors E_1dx_1, E_2dx_2, E_3dx_3 along the coordinate axes, is

$$dW_{123}=E_1E_2E_3dw=iE_{45}dw, \tag{4·63}$$

where $dw=dx_1dx_2dx_3$.

If T is any space vector

$$TdW_{123}=iTE_{45}dw=Sdw, \tag{4·64}$$

where

$$S=iTE_{45}. \tag{4·65}$$

Then S is the three-dimensional vector density which corresponds to the vector T. We shall later meet with S in another connection, in which it is called the *strain vector* associated with the space vector T.

By (4·65) we can find the components s_μ of S in terms of the components t_μ of T. The following is the complete scheme of relation:

$$\begin{aligned} &s_{23}, s_{31}, s_{12}, s_4, s_5, s_{45}=-t_1, -t_2, -t_3, -it_5, -it_4, -it_{16}, \\ &s_1, s_2, s_3, s_{15}, s_{25}, s_{35}, s_{14}, s_{24}, s_{34}, s_{16} \\ &\quad=-t_{23}, -t_{31}, -t_{12}, it_{14}, it_{24}, it_{34}, -it_{15}, -it_{25}, -it_{35}, -t_{45}. \end{aligned} \tag{4·66}$$

The components of S associated with real matrices are given in the first line and those with imaginary matrices in the second line. It will be seen that the latter correspond to the 10-vector part of T.

4·7. Wave Functions.

Consider a wave vector ψ which is a function of a complete space vector T. That is to say, to every one of a continuous set of space vectors T there corresponds a wave vector which we denote by $\psi = f(T)$, or by $f(t_1, t_2, \ldots t_{16})$, where $t_1, t_2, \ldots$ are the components of T.

It is usual to distinguish certain of the variables t_μ as *coordinates*, the others being *parameters* of the wave function f. Denoting the coordinates by x_μ, and the parameters collectively by a, the notation is changed to $\psi_a = f_a(x_\mu)$. Thus the original wave function is treated as a continuous set of wave functions ψ_a distinguished by parameters a, each of which covers the domain of coordinates x_μ.

In most practical applications the domain of coordinates is space-time; and a wave function has the form $\psi = f(x_1, x_2, x_3, x_4)$, or $f(X)$, where $X = E_1 x_1 + E_2 x_2 + E_3 x_3 + E_4 x_4$. The function then specifies a wave vector field in space-time.

The function f is necessarily *double-valued*. To show this, let

$$\psi = f(x_1, x_2, x_3, x_4) = F(r, \theta, \phi, x_4) \tag{4·71}$$

i.e a wave-vector field, r, θ, ϕ being polar coordinates. If we apply a tensor transformation $q = e^{\frac{1}{2}E_{12}\alpha}$ to the frame, we refer the vector field to a new coordinate system. By (4·15) the effect of the transformation on the co-ordinate system is that the point whose azimuth was ϕ in the old system has an azimuth $\phi + \alpha$ in the new system. The new coordinates are therefore

$$r' = r, \quad \theta' = \theta, \quad \phi' = \phi + \alpha, \quad x_4' = x_4. \tag{4·72}$$

The transformation of ψ is $$\psi' = e^{\frac{1}{2}E_{12}\alpha}\psi. \tag{4·73}$$

Let the vector field in the new coordinate system be

$$\begin{aligned}\psi' &= f'(x_1', x_2', x_3', x_4') = F'(r', \theta', \phi', x_4') \\ &= F'(r, \theta, \phi + \alpha, x_4).\end{aligned} \tag{4·74}$$

Then, by (4·73), $$F'(r, \theta, \phi + \alpha, x_4) = e^{\frac{1}{2}E_{12}\alpha}F(r, \theta, \phi, x_4). \tag{4·75}$$

Now let $\alpha = 2\pi$. Since $e^{E_{12}\pi} = -1$,

$$F'(r, \theta, \phi + 2\pi, x_4) = -F(r, \theta, \phi, x_4), \tag{4·76}$$

or in rectangular coordinates

$$f'(x_1, x_2, x_3, x_4) = -f(x_1, x_2, x_3, x_4). \tag{4·77}$$

But, since the axes after rotation through 2π are the same as they were originally, $\psi = f(x_1, x_2, x_3, x_4)$ and $\psi = f'(x_1, x_2, x_3, x_4)$ are both expressions for the wave-vector field in the original coordinate system. Thus ψ is a double-valued function $\pm f$ of rectangular coordinates.

Since this result is of vital importance, we must try to remove any doubt as to its meaning. Let Σ be a particular frame of rectangular coordinates and time, and let P be a particular point in that frame.

Suppose first that ψ is single-valued; so that, referred to the frame Σ and at the point P, it has the value ψ_0 and no other; in particular $\psi \neq -\psi_0$. When we vary the frame or the point considered, ψ changes; but if after such variation we return to the frame Σ and the point P, ψ must return to ψ_0. For example, if we rotate the frame in any plane in space, after a rotation 2π we come back to the frame Σ again, so that ψ is again ψ_0. But if ψ is a wave vector, the law of transformation of wave vectors requires that, when the space-time axes are rotated through 2π, ψ shall change continuously from ψ_0 to $-\psi_0$. Therefore the single-valued ψ which we have been considering cannot satisfy the transformation law of a wave vector.

The double-valuedness of f becomes of practical importance when (as usual) the vector field is defined over a region which includes all azimuths ϕ. For then, on following the point P round a circuit back to the initial azimuth, we may find ourselves on the opposite branch of f from that on which we started. The return may be either to the opposite branch or the same branch; either type of connection satisfies the condition of continuity of ψ in the region over which it is defined.

We note accordingly (for future reference) that a symbol ψ_α ($\alpha = 1, 2, 3, 4$), defined to be a single-valued function of rectangular coordinates over a domain which includes or encircles the origin, cannot be a wave vector. By "encircles the origin" we mean that a circle having the origin as centre can be drawn in it.

CHAPTER V

THE SIMPLE WAVE EQUATION

5·1. Invariant Equations.

In order that the laws of physics may be independent of the choice of frame (among the equivalent orthogonal frames) they must be expressible as tensor equations. In wave-tensor calculus, the simplest non-trivial tensor equation is of the form

$$H\psi = 0, \tag{5·11}$$

where H is a mixed wave tensor and ψ a covariant wave vector. Then (5·11) is a vector equation $H_\alpha{}^\beta \psi_\beta = 0$ equivalent to four algebraic equations.

It may be anticipated that the simplest, and presumably the most fundamental, laws of physics will have this form. Alternatively, regarding (5·11) as a definition rather than a law, it is an appropriate means of introducing a wave vector ψ and relating it to the ordinary space vectors of physics. For we have seen (§ 4·4) that a mixed wave tensor H is constituted of space tensors which will presumably be recognised as such in our practical observations; but there is no such "projection" of ψ into a space-time representation, and its connection with the ordinary space tensors of physics can only be expressed indirectly by an equation such as (5·11).

It is appropriate to introduce simultaneously a contravariant wave vector χ^*, satisfying

$$\chi^* H = 0. \tag{5·12}$$

We may expect that ψ and χ^* will occur symmetrically in physical theory.

Thus far our argument has been that if ever nature condescends to simplicity, equations of the types (5·11) and (5·12) will figure in her scheme.

Before the birth of wave mechanics the systematised part of physics was wholly described by space vectors and tensors. Wave mechanics introduced a new kind of entity ψ. It was introduced in the way here proposed by a "wave equation" in which the coefficients were the ordinary space-tensor quantities of physics. The original ψ of Schrödinger did not satisfy the relativity requirements of atomic physics; but in 1928 Dirac introduced a ψ with four components, which satisfied an equation invariant for the six relativity rotations of space-time, although the invariance was not of a kind contemplated in the usual tensor calculus. Our anticipated fundamental equation turns out to be a form of Dirac's equation.†

† The form given by Dirac (*Quantum Mechanics*, 2nd ed., p. 255, equations (9) and (10)) is the equivalent strain vector equation, which we shall obtain in (7·73). Dirac further postulates, as a "reality condition", that the two wave vectors are conjugate complex quantities. Our reality conditions are determined directly from relativity principles in Chapter VI, and do not impose this restriction.

The mixed wave tensor H used in Dirac's equation is limited to five components. One component has dropped out through special choice of axes, permissible when as usual the region contemplated is small enough to be treated as flat. Apart from this the truncation is significant, because the general mixed wave tensor T cannot be reduced to Dirac's special form H by any choice of axes. We shall account for this limitation in § 5·4.

We shall derive Dirac's equation according to the principles which we are developing in § 5·4; but we shall first examine its elementary properties—looking ahead to see the theory which we are about to meet.

Dirac's equations are $$H\psi = 0, \quad \chi^* H = 0, \tag{5·13}$$

where $$H = E_1 p_1 + E_2 p_2 + E_3 p_3 + E_4 p_4 - m \tag{5·14}$$

and E_1, E_2, E_3, E_4 constitute a tetrad. We call H the *hamiltonian*.† Since H is required to be a mixed wave tensor, its components form space tensors in accordance with the specification in (4·42), namely (p_1, p_2, p_3, p_4) is a space vector and the quarterspur $-m$ is an invariant.

Conversely if, following Dirac, we construct a hamiltonian H out of a "momentum vector" (p_1, p_2, p_3, p_4) and an invariant mass m by the formula (5·14), H will be a mixed wave tensor; and therefore the wave equations $H\psi = 0$, $\chi^* H = 0$ will be tensor equations which continue to be satisfied when any of the six relativity transformations of space-time are applied.

Thus the invariance of Dirac's equation for relativity transformations, which was a novel kind of invariance from the point of view of ordinary tensor calculus, is an elementary consequence of wave-tensor calculus.

5·2. Properties of Dirac's Equation.

The equation $H\psi = 0$ shows that H has an eigenvalue 0, and hence that it is singular (§ 3·7 (*b*)). It has a pseudo-reciprocal

$$H' = E_1 p_1 + E_2 p_2 + E_3 p_3 + E_4 p_4 + m. \tag{5·21}$$

For, multiplying by (5·14),

$$HH' = -p_1^2 - p_2^2 - p_3^2 - p_4^2 - m^2. \tag{5·22}$$

The product contains no non-algebraic terms, and therefore vanishes by (2·66).

For a physically real momentum vector (p_1, p_2, p_3, p_4), p_1, p_2, p_3 are real and p_4 is imaginary. Let

$$p_4 = ip_0, \tag{5·23}$$

† In classical theory the hamiltonian is the expression for the energy $(-ip_4)$ in terms of the momenta p_1, p_2, p_3 and coordinates. If, following the usual relativistic view, time is treated on the same footing with the other coordinates, the hamiltonian is correspondingly defined as the expression for the proper energy m in terms of p_1, p_2, p_3, p_4 and the coordinates. This would make the hamiltonian strictly $H + m$, but we shall use the term without regard to an additive constant.

so that p_0 is the real time component, i.e. the energy or mass. The vanishing of (5·22) gives

$$m^2 = p_0^2 - p_1^2 - p_2^2 - p_3^2. \tag{5·24}$$

This identifies $\pm m$ with the proper energy (or proper mass) corresponding to the momentum vector. Thus m must be real.

By (5·24)
$$p_0 = (m^2 + p_1^2 + p_2^2 + p_3^2)^{\frac{1}{2}}. \tag{5·251}$$

Usually p_1, p_2, p_3 are small compared with m, and we then have the "classical" approximation for the energy

$$p_0 = m + (p_1^2 + p_2^2 + p_3^2)/2m. \tag{5·252}$$

The general solution of the wave equations is found as follows. Let ϕ, ω^* be arbitrary four-valued quantities. Since $HH' = H'H = 0$, we have

$$H(H'\phi) = 0, \quad (\omega^* H')H = 0.$$

Thus
$$\psi = H'\phi, \quad \chi^* = \omega^* H' \tag{5·26}$$

are solutions of (5·13). Inserting row-and-column suffixes, these become

$$\left.\begin{aligned} \psi_\alpha &= H_{\alpha 1}'\phi_1 + H_{\alpha 2}'\phi_2 + H_{\alpha 3}'\phi_3 + H_{\alpha 4}'\phi_4, \\ \chi_\alpha &= \omega_1 H_{1\alpha}' + \omega_2 H_{2\alpha}' + \omega_3 H_{3\alpha}' + \omega_4 H_{4\alpha}'. \end{aligned}\right\} \tag{5·27}$$

Since ϕ_1, ϕ_2, ϕ_3, ϕ_4 are arbitrary coefficients, the general value of ψ_α is a linear combination of four elementary solutions $H_{\alpha 1}', H_{\alpha 2}', H_{\alpha 3}', H_{\alpha 4}'$, i.e. the four rows of the matrix H'. Similarly the four columns of H' are the elementary solutions for χ_α. Since H' is singular, its determinant vanishes, and therefore only three of the rows and three of the columns are linearly independent. There are therefore not more than three independent solutions in each case.

For example, use the matrix representation (3·61). The special form (5·21) for H' gives

$$-iH' = \begin{matrix} p_2 - im & p_1 + p_0 & 0 & p_3 \\ p_1 - p_0 & -p_2 - im & -p_3 & 0 \\ 0 & -p_3 & p_2 - im & p_1 - p_0 \\ p_3 & 0 & p_1 + p_0 & -p_2 - im. \end{matrix} \tag{5·28}$$

Any row gives a solution for ψ and any column a solution for χ^*. It is not necessary to choose the solutions in a corresponding way.† For example, we might take

$$\psi = (p_2 - im, p_1 + p_0, 0, p_3), \quad \chi^* = (0, -p_3, p_2 - im, p_1 + p_0)$$

as the pair of wave vectors constituting a solution of the wave equation.

† Our treatment here differs fundamentally from that of Dirac. See footnote, p. 62; also § 8·6.

5·3. The Stream Vector.

Let ψ, χ^* be solutions of the equations $H\psi=0$, $\chi^*H=0$; and let

$$\psi\chi^*=J=\Sigma j_\mu E_\mu. \tag{5·31}$$

Multiply the wave equations by initial χ^*E_{12} and final $E_{12}\psi$, respectively; we obtain

$$\chi^*(E_2p_1-E_1p_2+E_{12}E_3p_3+E_{12}E_4p_4-E_{12}m)\psi=0,$$
$$\chi^*(-E_2p_1+E_1p_2+E_3E_{12}p_3+E_4E_{12}p_4-E_{12}m)\psi=0.$$

Hence, subtracting,

$$2\chi^*(E_2p_1-E_1p_2)\psi=0,$$

so that, by (3·37),

$$j_2p_1-p_1j_2=0. \tag{5·321}$$

Again, multiplying by initial χ^*E_{15} and final $E_{15}\psi$, and subtracting,

$$\begin{aligned}0&=\chi^*(E_{15}H-HE_{15})\psi\\&=2\chi^*(E_5p_1)\psi=-8j_5p_1,\end{aligned} \tag{5·322}$$

so that $j_5=0$. Also multiplying by initial χ^*E_1 and final $E_1\psi$, and adding,

$$\begin{aligned}0&=\chi^*(E_1H+HE_1)\psi\\&=2\chi^*(-p_1-E_1m)\psi\\&=-8(ip_1j_{16}-j_1m).\end{aligned} \tag{5·323}$$

The results (5·321), (5·322), (5·323) give

$$\frac{j_1}{p_1}=\frac{j_2}{p_2}=\frac{j_3}{p_3}=\frac{j_4}{p_4}=\frac{j_5}{0}=\frac{ij_{16}}{m}. \tag{5·33}$$

The wave equation can therefore be written in the equivalent form

$$(E_1j_1+E_2j_2+E_3j_3+E_4j_4-E_{16}j_{16})\psi=0. \tag{5·34}$$

We call (j_1, j_2, j_3, j_4) the *stream vector*. The whole set of sixteen j_μ is the *complete stream vector*. We have here proved that the stream vector is equal to the momentum vector except for a numerical factor. Multiplying the complete stream vector by the same factor we obtain the *complete momentum vector*.

We regard this correspondence of the stream vector and momentum vector as a coalescence which occurs in the peculiarly simple system here studied. In more general physical systems they are not so closely connected. According to the definition of the momentum vector usually adopted in quantum theory and reached later in this book, the components p_μ are not necessarily algebraic quantities; they may be matrices or general symbols. On the other hand the components j_μ of the stream vector are necessarily algebraic quantities.

The result (5·34) leads us to a new view of the wave equation. Consider a physical system described by a *pure* (i.e. factorisable) wave tensor J, or by the equivalent set of space tensors. We are not given the whole set of space tensors, but only one of the space vectors (j_1, j_2, j_3, j_4) together with the two invariants $j_5\,(=0)$ and j_{16}. We cannot therefore determine definitely

the factors of J; but our data are sufficient to limit them to certain possibilities, viz. they must be solutions of (5·34) and of the corresponding equation for χ^*.

The wave equation is therefore an equation for determining the possible factors of a wave tensor, which is only partly known.

At this point we must try to make clear a difference in our attitude towards wave mechanics from that which appears to be usual among quantum physicists. It will probably be agreed that wave mechanics is a *method* of analysis, not a *theory* of phenomena. The ψ waves have no objective existence; we invent them as required in solving our problems. In the present treatment we have found that any space vector can be expressed as a wave tensor. Ordinarily it is not a pure wave tensor; but it can be represented as a sum of pure wave tensors, which are then resolved into their wave-vector factors. If the space vector is a function of the coordinates, the wave-vector factors become "wave functions". In this way wave functions appear in connection with any characteristic of a system which is described by space vectors. It is therefore ambiguous to speak of the wave functions of a system; we should rather speak of the wave functions associated with some specified tensor of the system. Reference to the wave function or the wave equation of a system leaves us in the same state of conjecture as if reference were made to "the tensor of the hydrogen atom" or "the equation of the sun".

I am not cavilling at expressions, whose meaning is doubtless made plain either by the context or by custom. My point is that when wave analysis is our standard procedure—when the ordinary tensor calculus is replaced by wave-tensor calculus—we shall introduce new wave functions as casually as we introduce new tensors. The domain of physics treated in this book is for the most part different from that which has occupied the attention of writers on pure quantum theory. Sometimes our wave functions will coincide with theirs; sometimes they will differ. We find it well to maintain a certain amount of contact in order to utilise well-known results; but in principle we do not bind ourselves to use the wave functions that the quantum physicists have discussed. Remembering that the introduction of wave functions is merely a factorisation, we must obviously retain freedom to employ factorisation whenever it is useful.

The reader must therefore be prepared to find here a greater elasticity in the definition and use of wave functions than he has been accustomed to.

5·4. The Wave Equation as an Identity.

If we represent E_1, E_2, E_3, E_4, E_5 by the special pentad of matrices (3·27), it is not difficult to prove by straightforward verification that

$$\sum_{\mu=5} (E_\mu\psi)_\alpha (E_\mu\psi)_\beta - (E_{16}\psi)_\alpha (E_{16}\psi)_\beta \equiv 0, \qquad (5{\cdot}41)$$

where ψ is any four-valued quantity. Here, as usual, $E_\mu\psi$ is the four-valued quantity formed by chain multiplication, and $(E_\mu\psi)_\alpha$ is one of its four components, i.e. $(E_\mu\psi)_\alpha=(E_\mu)_{\alpha\beta}\psi_\beta$.

Any other pentad E_μ' is obtained by a transformation $E_\mu'=qE_\mu q^{-1}$. Let $\psi'=q\psi$; then $E_\mu'\psi'=qE_\mu\psi$, so that

$$(E_\mu'\psi')_\sigma(E_\mu'\psi')_\tau=(qE_\mu\psi)_\sigma(qE_\mu\psi)_\tau$$
$$=q_{\sigma\alpha}(E_\mu\psi)_\alpha q_{\tau\beta}(E_\mu\psi)_\beta.$$

Hence, multiplying (5·41) by $q_{\sigma\alpha}q_{\tau\beta}$, we have

$$\sum_{\mu=1}^{\mu=5}(E_\mu'\psi')_\sigma(E_\mu'\psi')_\tau-(E_{16}\psi')_\sigma(E_{16}\psi')_\tau=0. \qquad (5\cdot42)$$

We can choose ψ' arbitrarily; because (since q is not singular) the corresponding $\psi=q^{-1}\psi'$ can be employed in (5·41).

Thus the identity (5·41), verified for a particular pentad, is true for any pentad of matrices whatsoever.

Let χ^* be another arbitrary four-valued quantity. Multiply (5·41) by initial χ_α (inner multiplication). We have by (3·37)

$$\chi_\alpha(E_\mu\psi)_\alpha=\chi^*E_\mu\psi=-4j_\mu,$$

where $\psi\chi^*=J=\Sigma j_\mu E_\mu$. The result is therefore

$$\sum_1^5 j_\mu(E_\mu\psi)_\beta-j_{16}(E_{16}\psi)_\beta=0$$

or
$$(E_1j_1+E_2j_2+E_3j_3+E_4j_4+E_5j_5-E_{16}j_{16})\psi=0. \qquad (5\cdot43)$$

We can show similarly that

$$\chi^*(E_1j_1+E_2j_2+E_3j_3+E_4j_4+E_5j_5-E_{16}j_{16})=0. \qquad (5\cdot44)$$

Except that the term in j_5 is included, these are the wave equations as given in (5·34). They are here obtained as an identity satisfied by any two wave vectors ψ, χ^* and their outer product J.

We see that Dirac was right in restricting his hamiltonian to the above terms, instead of employing the sixteen terms of a complete space vector. By omitting j_5 he restricts the equation to stream vectors which have zero component normal to space-time; otherwise his equation is a perfectly general one satisfied by the factors of any pure wave tensor. The hamiltonian H is part of the complete stream vector J, except that the sign of j_{16} is reversed. It must not be supposed that the components of J (other than j_5) which do not appear in H are zero.

On the other hand Dirac's postulate that ψ and χ^* (or rather a quantity ϕ^* easily derived from χ^*) are conjugate complex quantities would restrict

J to very special forms. The restriction has to do with certain special applications, and is inappropriate in general theory.

Multiply (5·43) by initial χ^* and apply (3·37). We obtain

$$j_1^2+j_2^2+j_3^2+j_4^2+j_5^2-j_{16}^2=0. \tag{5·45}$$

Again, multiply (5·43) by initial $\chi^* E_5$ and apply (3·37). We obtain

$$j_1 j_{15}+j_2 j_{25}+j_3 j_{35}+j_4 j_{45}=0. \tag{5·46}$$

These, and the corresponding equations obtained by substituting other pentads, are relations satisfied identically by the components of a pure wave tensor. There are, of course, no such relations between the components of a general wave tensor T which is not stated to be factorisable.

Let $$P_\alpha=\sum_{\mu=0}^{\mu=5} j_{\alpha\mu}E_{\alpha\mu} \quad (\alpha=0,\ 1,\ 2,\ 3,\ 4,\ 5;\ \mu\neq\alpha). \tag{5·47}$$

For fixed α, the matrices $E_{\alpha\mu}$ form a pentad. We therefore call P_α a *pentadic part* of J. There are six pentadic parts which overlap, so that

$$\tfrac{1}{2}\Sigma_\alpha P_\alpha+\mathrm{qs}\,J=J. \tag{5·48}$$

The pentad which we have been using corresponds to $\alpha=0$, and the wave equation (5·43) can be written

$$(P_0-ij_{16})\psi=0.$$

But since the proof holds for any pentad, we have more generally

$$(P_\alpha-ij_{16})\psi=0 \tag{5·491}$$

or equivalently

$$P_\alpha\psi=(\mathrm{qs}\,J)\psi \tag{5·492}$$

for all six values of α.

From the present standpoint the use of the wave equation is to determine the factors of a pure wave tensor J. It seems to be generally true that in physics we determine a factor ψ, not because its value is of particular importance to us, but because that happens to be the most convenient way of ascertaining that a factor exists, i.e. that J is pure. For example, the wave function ψ of a hydrogen atom is investigated primarily because the mere existence of such a function imposes certain conditions on the hamiltonian (which is part of J), and these conditions determine the energy levels of the atom. It would probably be difficult to solve the more complex problems of quantum theory without evaluating ψ; but since the observables of physics are always space tensors and therefore derived from wave tensors of the second (or higher) rank, the wave vector factors must ultimately be recombined.

In the present case (which is perhaps too elementary to be typical) the conditions for purity of J are expressed directly by the equations (5·45) and (5·46), and there is no need to evaluate the factors

5·5. Standard Forms of Pure Wave Tensors.

Equation (5·492) asserts:

A factor of J is an eigensymbol of every pentadic part of J, and the eigenvalue of a pentadic part is qs J. (5·51)

Consider an antitriad $E_{\mu\nu}, E_{\sigma\tau}, E_{\lambda\rho}$ (2·36). Each pentad contains one and only one member of an antitriad. Hence, in the expression $(E_{\mu\nu}+E_{\sigma\tau}+E_{\lambda\rho})m$, each term is a pentadic part and has eigenvalues $\pm im$. Accordingly the form

$$J=(\pm E_{\mu\nu}\pm E_{\sigma\tau}\pm E_{\lambda\rho}+E_{16})m \tag{5·52}$$

will, if the signs are properly chosen, satisfy the condition (5·51) that the eigenvalue of every pentadic part is equal to the quarterspur. It turns out that four of the combinations of sign make J factorisable, and four do not. For example, take the + sign for the first two terms; then by (5·492)

$$E_{\mu\nu}\psi=i\psi, \quad E_{\sigma\tau}\psi=i\psi. \tag{5·53}$$

Hence, if $\mu, \nu, \sigma, \tau, \lambda, \rho$ is an even permutation of 0, 1, 2, 3, 4, 5,

$$iE_{\lambda\rho}\psi=E_{\mu\nu}E_{\sigma\tau}\psi=i^2\psi,$$

so that the + sign must also be taken for the third term in order to satisfy (5·492). It is then easy to verify that J is factorisable by working out the factors in a particular matrix representation, or by testing it for idempotency according to the theory given in the next section.

Accordingly our result is that

$$J=(E_{\mu\nu}+E_{\sigma\tau}+E_{\lambda\rho}+E_{16})m \tag{5·54}$$

is a pure matrix if $\mu, \nu, \sigma, \tau, \lambda, \rho$ is an even permutation. Any two of the first three terms can be given negative sign, since this is equivalent to reversing the order of their suffixes and the permutation remains even.

Any non-degenerate pure wave tensor can be reduced to the standard form (5·54) by a relativity transformation $J'=qJq'$. We first make a transformation so that one of the components, say j_5, becomes zero. Then, by (5·45),

$$j_1^2+j_2^2+j_3^2+j_4^2=j_{16}^2, \quad j_{15}^2+j_{25}^2+j_{35}^2+j_{45}^2=j_{16}^2.$$

Since J is non-degenerate, $j_{16}\neq 0$. Hence the vectors (j_1, j_2, j_3, j_4) and $(j_{15}, j_{25}, j_{35}, j_{45})$ have the same non-zero length, and by (5·46) they are at right angles. We can therefore choose two of the axes in four dimensions to coincide with them; we then have

$$j_1=j_{25}=j_{16}, \quad j_2=j_3=j_4=j_{15}=j_{35}=j_{45}=j_5=0. \tag{5·551}$$

Applying (5·492) with $\alpha=0, 5, 3$, we now have

$$E_1j_1\psi=ij_{16}\psi, \quad E_{25}j_{25}\psi=ij_{16}\psi, \quad (E_{31}j_{31}+E_{32}j_{32}+E_{34}j_{34})\psi=ij_{16}\psi. \tag{5·552}$$

Thus ψ is an eigensymbol of the two commuting symbols E_1, E_{25} and therefore of their product iE_{34}. It is therefore an eigensymbol of $(E_{31}j_{31}+E_{32}j_{32})$;

and the eigenvalue must be zero because $(E_{31}j_{31}+E_{32}j_{32})$ anticommutes with E_{34}. Hence the third equation of (5·552) breaks up into

$$(E_{31}j_{31}+E_{32}j_{32})\psi=0, \quad E_{34}j_{34}\psi=ij_{16}\psi,$$

so that

$$j_{31}=\pm ij_{32}, \quad j_{34}=\pm j_{16}.$$

Hence $j_{31}(E_{31}\pm iE_{32})\psi=0$; or, multiplying by E_{31}, $j_{31}(-1\pm iE_{12})\psi=0$. Now ψ cannot be an eigensymbol of E_{12}, because it is an eigensymbol of E_1 which anticommutes with E_{12}. Hence

$$j_{31}=0, \quad j_{32}=0.$$

Similarly we find $j_{41}=0, j_{42}=0$.

The pentad $\alpha=1$ then gives $(E_1j_1+E_{12}j_{12})\psi=ij_{16}\psi$. Hence, by the first equation of (5·552), $j_{12}=0$. All the components are now accounted for, and J reduces to $(E_1+E_{25}+E_{34}+E_{16})j_{16}$, which is of the required standard form.

To obtain a standard form for a *degenerate* pure wave tensor we proceed as follows. Let $J=\psi\chi^*$ be a degenerate pure wave tensor ($j_{16}=0$), and let j_ϖ be a component which does not vanish. Then $E_\varpi J$ is a pure wave tensor, since it has factors $E_\varpi\psi$ and χ^*; and it is non-degenerate since its quarterspur is $E_\varpi(E_\varpi j_\varpi)=-j_\varpi$. Hence $E_\varpi J$ can be reduced to the form (5·54) by a relativity transformation. We take therefore

$$E_\varpi J=(E_{\mu\nu}+E_{\sigma\tau}+E_{\lambda\rho}+E_{16})m.$$

E_ϖ cannot be $E_{\mu\nu}$, $E_{\sigma\tau}$, $E_{\lambda\rho}$ or E_{16}, since J would then be non-degenerate; but it can be any other E-symbol. Taking $E_\varpi=E_{\mu\sigma}$, we obtain

$$J=(E_{\nu\sigma}+E_{\mu\tau}+iE_{\nu\tau}+iE_{\sigma\mu})m. \tag{5·56}$$

This is the standard form for a degenerate pure wave tensor.

5·6. Idempotency.

A symbol J is said to be *idempotent* if $J^2=J$.

To *normalise* an E-number we multiply it by an algebraic factor so as to make the quarterspur $\frac{1}{4}$. If it is represented as a matrix (so that a spur exists) we normalise it by making the spur 1. It is, of course, impossible to normalise a degenerate E-number.

We shall show that *a necessary and sufficient condition for a non-degenerate matrix to be pure is that it shall be idempotent when normalised.*

(5·61)

Let $J=\psi\chi^*$ be a normalised matrix so that

$$\text{spur } J=\chi^*\psi=1.$$

Then

$$J^2=\psi\chi^*\psi\chi^*=\psi.1.\chi^*=J,$$

so that the condition is necessary. To prove that it is sufficient, let T be a matrix satisfying

$$T^2=T, \quad \text{spur } T=1. \tag{5·62}$$

Any matrix can be expressed as the sum of a number of vector products; therefore let

$$T=\psi_a\chi_a{}^*+\psi_b\chi_b{}^*+\psi_c\chi_c{}^*+\ldots. \tag{5·63}$$

Here the suffixes $a, b, c, \ldots$ distinguish different vectors, the row-and-column suffixes being omitted as usual. We write A_{ab} for the scalar product $\chi_a{}^*\psi_b$, so that the product $\psi_a\chi_a{}^*\psi_b\chi_b{}^*$ reduces to $A_{ab}\psi_a\chi_b{}^*$. Then

$$T^2-T=(A_{aa}-1)\psi_a\chi_a{}^*+A_{ab}\psi_a\chi_b{}^*+A_{ba}\psi_b\chi_a{}^*+\ldots=0. \tag{5·64}$$

Corresponding to the four suffixes of χ this gives four linear equations satisfied by the vectors $\psi_a, \psi_b, \psi_c, \ldots$. Using any one of these equations to give the value of ψ_a in terms of the other ψ's, we can eliminate ψ_a in (5·63) and so reduce by one the number of vector products on the right-hand side of (5·63). Repeating the process, we reduce the number of vector products one by one.

The procedure fails if the coefficients of ψ_a vanish in all four equations, i.e. if

$$(A_{aa}-1)\chi_a{}^*+A_{ab}\chi_b{}^*+A_{ac}\chi_c{}^*+\ldots=0. \tag{5·65}$$

But we can then use (5·65) to eliminate $\chi_a{}^*$ in (5·63), and the number of vector products is again reduced by one.

The reduction can be continued so long as there are any non-vanishing coefficients in (5·64). When all the coefficients vanish so that

$$A_{aa}=A_{bb}=A_{cc}=\ldots=1, \quad A_{ab}=A_{ba}=\ldots=0,$$

no further reduction is possible. We then have, by (5·63),

$$\begin{aligned}\text{spur } T&=\text{spur }\psi_a\chi_a{}^*+\text{spur }\psi_b\chi_b{}^*+\text{spur }\psi_c\chi_c{}^*+\ldots\\&=A_{aa}+A_{bb}+A_{cc}+\ldots\\&=1+1+1+\ldots.\end{aligned}$$

But spur $T=1$, so that there can be only one term on the right-hand side. That is to say, T is the product of two vectors.

A pure matrix is necessarily singular. This follows from § 3·7 (*b*), since the idempotent condition, $J^2-J=0$, gives eigenvalues 0 and 1. A singular matrix is not necessarily pure.

If the square of a matrix is -1, the question sometimes arises whether it has to fulfil any other condition in order that it may be a member of a complete orthogonal set. The most commonly occurring combinations whose squares are algebraic are triadic and pentadic expressions; these can serve as individual members of a new complete set. But it has been pointed out by D. E. Littlewood† that we can also form combinations of antiperpendicular matrices whose squares are algebraic, and these cannot be members of a complete set. We find by direct multiplication that the square of

$$\tfrac{1}{2}(E_{\mu\nu}+E_{\sigma\tau}-E_{\lambda\rho}+E_{16}) \tag{5·66}$$

† *Journ. Lond. Math. Soc.* **9**, 41 (1934).

is -1. Of the eight possible combinations of sign in an antitetrad, four yield factorisable matrices, as we have seen; the other four give matrices whose squares are algebraic.

The quarterspur of (5·66) is $\frac{1}{2}i$, whereas the quarterspur of E_μ is 0 or i. Since the quarterspur is invariant for the transformation $F_\mu = qE_\mu q'$, there can be no such transformation connecting (5·66) and E_μ. Therefore (5·66) cannot be a member of a complete set. We shall call an expression of the form (5·66), or reducible to it by a relativity transformation, a *compact* E-number.

I have not as yet found any physical application for compact E-numbers; but perhaps others will be more successful. They surely must have an importance of some kind, possibly in the theory of radiation or even in the theory of the nucleus—subjects which we do not seriously attempt to treat in this book.

5·7. Spectral Sets.

We suggested in § 5·4 that the wave vector ψ was investigated in physics, not for its own sake, but because the existence of factors imposes certain invariant conditions on the stream vector and on the hamiltonian which forms part of it. We may now go a step further, and say that the condition which it is sought to impose is that of idempotency. Those familiar with the Group Theory of wave mechanics will recall the fundamental part played by idempotent operators in selecting the "pure" states of a statistical *ensemble*.

Consider the wave tensors

$$\left.\begin{aligned} J_a &= -\tfrac{1}{4}i \quad (E_{\mu\nu} + E_{\sigma\tau} + E_{\lambda\rho} + E_{16}), \\ J_b &= -\tfrac{1}{4}i\,(-E_{\mu\nu} - E_{\sigma\tau} + E_{\lambda\rho} + E_{16}), \\ J_c &= -\tfrac{1}{4}i\,(-E_{\mu\nu} + E_{\sigma\tau} - E_{\lambda\rho} + E_{16}), \\ J_d &= -\tfrac{1}{4}i \quad (E_{\mu\nu} - E_{\sigma\tau} - E_{\lambda\rho} + E_{16}). \end{aligned}\right\} \tag{5·71}$$

We have found in (5·54) that these are pure. Since the quarterspur is $-\frac{1}{4}iE_{16} = \frac{1}{4}$, they are normalised. Hence they are idempotent, as can be verified by direct multiplication. We can also verify that their products are zero. They accordingly satisfy

$$J_a^2 = J_a, \quad J_a J_b = 0, \quad J_a + J_b + J_c + J_d = 1. \tag{5·72}$$

A set of operators satisfying the conditions (5·72) is called a *spectral set*. Here the set consists of four operators only. The more familiar examples of spectral sets in physics include an infinite number of operators. For example, let J_λ denote the operation of selecting light of wave length λ from a source of light represented by ψ; thus the light of wave length λ existing in the source is represented by $J_\lambda\psi$. If we repeat the selective operation J_λ on $J_\lambda\psi$, it makes no difference; hence $J_\lambda^2 = J_\lambda$. The symbol $J_{\lambda'}J_\lambda$ denotes the operation of selecting wave length λ' out of light already selected as being

of wave length λ; the result is obviously zero. Further, selecting every wave length in turn and adding the results, we reproduce the original source of light; hence $\Sigma_\lambda J_\lambda$ is equal to the "identical operator" 1. The selective operators of spectral analysis therefore fulfil the equations (5·72), which ensure that they are idempotent, non-overlapping and exhaustive.

As G. Temple has pointed out† it is equations of the form (5·72) which directly embody the physical conception of a "pure constituent". The mathematically convenient criterion of purity, namely factorisability of the operator in matrix representation, should be regarded as derived from (5·72) rather than *vice versa*.

This suggests a new approach to the theory of the representation of phenomena by E-symbols. We can regard the matrices E_μ as introduced by a spectral analysis of entities represented by algebraic numbers (in particular, probability distributions or densities) into four pure constituents given by (5·71). This point of view is developed in § 13·6.

5·8. The Complete Stream Vector of a Particle.

Consider a particle in spherical space-time. A classical particle is described by two 4-vectors, namely a position vector and a velocity vector. In five-dimensional representation the position vector is the radius of space-time which passes through the particle; the velocity vector is at right angles to it and lies in the four-dimensional hypersphere.

If we take axes such that the position vector is in the E_5 direction and the velocity vector is in the E_4 direction, the two vectors reduce to single components j_5 and j_4. Let us treat them as components of a single wave tensor. There is, of course, no compulsion to combine them; there is no unique definition of *the* wave tensor of a particle, any more than in ordinary relativity theory there is a unique definition of *the* tensor of a particle; and it would be legitimate to investigate a tensor representing position only or velocity only, if desired. But we shall try to find a tensor, called the *complete stream vector*, which comprises both.

If the complete stream vector is pure, it must have two more components besides j_4 and j_5. We may take it to be

$$(E_{41}+E_{50}+iE_{40}+iE_{15})\,\alpha, \tag{5·81}$$

which is of the standard form (5·56) with $\mu, \nu, \sigma, \tau = 5, 4, 1, 0$. The additional terms define an axis in the three-dimensional space, which is in some way characteristic of the particle. This axis, which in (5·81) is taken to be in the E_1 direction, is called the *spin axis*.

The question now arises whether in ascribing a complete stream vector J to the particle we should take (5·81) to be the actual vector J or the vector

† "The Physical Principles of the Quantum Theory", *Proc. Roy. Soc.* A, 138, 479.

density $J.iE_5$. This is answered by the Uncertainty Principle, which asserts that a particle cannot have exact position and exact velocity simultaneously. Thus our combination of a position vector and a velocity vector will not apply to a discrete particle, but describes an element of its probability distribution. We must therefore take (5·81) to be a vector density, so that

$$iJE_5 = (E_{41} + E_{50} + iE_{40} + iE_{15})\,\alpha. \tag{5·82}$$

From this we obtain $$J = (E_{23} + E_{16} + E_{45} + E_{01})\,\alpha, \tag{5·83}$$

which is of the standard form (5·54) for a non-degenerate pure tensor.

The direction of the spin axis is shown by the term E_{01}, or equivalently by the term E_{23} which gives the plane of the spin. The velocity vector, which by our choice of axes is in the time direction, is represented by the term E_{45}, this being the matrix of the rotation which would displace the particle in the time direction. Instead of a position vector, we have an invariant E_{16}. The "position" of the particle is therefore invariant for all relativity rotations; this is only possible if we represent the particle as an entity uniformly distributed throughout the hypersphere of space-time. This is in agreement with the uncertainty principle; for we have ascribed an exact velocity vector E_{45} to the particle, and therefore its position is entirely indeterminate.

The attempt to assign a combination of position vector and velocity vector to a particle breaks down, as the uncertainty principle foretells. The position vector E_{50} in (5·82) defines, not the position of the particle, but the position of an element of its probability distribution selected for consideration.

We define an *elementary particle* to be an entity whose characteristics are completely specified by a complete stream vector of the type (5·83), so that it can be represented by simple wave vectors ψ, χ^*. It has exact momentum but indeterminate position. This would perhaps more usually be called an elementary state of an elementary particle; and it is contemplated that a number of elementary states may be superposed—forming a wave packet which has approximate position and momentum. It is to be remembered, however, that the properties of observational significance are relations to other elementary particles or combinations of particles, and not the primitive relations to a symbolic frame summed up in (5·83). We must not be in too great a hurry to identify our formulae with those employed in the practical applications of quantum theory.

By (4·65) the three-dimensional vector density or strain vector corresponding to J is $S = iJE_{45}$. For the special wave tensor (5·83), we find

$$S = -J. \tag{5·84}$$

CHAPTER VI

REALITY CONDITIONS

6·1. Distinction between Space and Time.

In relativity theory the *interval* between two point-events is defined by its square ds^2. In Galilean coordinates

$$ds^2 = dt^2 - dx_1^2 - dx_2^2 - dx_3^2,$$

the velocity of light being taken to be unity as usual. By the use of anti-commuting symbols, the square root can be expressed in rational form; thus we may write

$$d\mathbf{s} = E_1 dx_1 + E_2 dx_2 + E_3 dx_3 + E_4 dx_4. \qquad (6{\cdot}11)$$

On squaring, the product terms cancel owing to the anticommutation, and we have

$$(d\mathbf{s})^2 = -dx_1^2 - dx_2^2 - dx_3^2 - dx_4^2 = ds^2$$

if $x_4 = it$. The algebraic square root $\pm ds$ is the eigenvalue of $d\mathbf{s}$. Since (6·11) is the space vector, or displacement, between the two points, our conclusion is that the interval is the eigenvalue of the displacement.

It is of fundamental importance that, since a pentad contains three imaginary and two real matrices, we cannot cover more than three dimensions with matrices of the same real or imaginary character. If then we use the imaginary matrices E_1, E_2, E_3 for the three similar space dimensions, we have to use a real matrix E_4 for the fourth dimension of the physical continuum of equivalent points. *Thus the distinctive character of the fourth dimension (time) is already foreshadowed in the constitution of the pentads.*

For real phenomena x_1, x_2, x_3 are real, and $x_4 (= it)$ is imaginary. Since E_1, E_2, E_3 are imaginary and E_4 is real, the vector interval $d\mathbf{s}$ is a wholly imaginary matrix.

A matrix which is wholly real or wholly imaginary will be called *monothetic*. Two matrices are *homothetic* if both are real or both imaginary, and *antithetic* if one is real and the other imaginary. The distinction between space and time is comprised in the statement that $d\mathbf{s}$ is a monothetic matrix.

In Schrödinger's wave mechanics certain "Hermitic conditions" were imposed in order that the mathematical expressions should correspond to real physical phenomena. It seems to have been generally assumed that the reality conditions of Dirac's wave mechanics must be of the same Hermitic form, although several writers have pointed out the difficulties arising from this assumption. In my own developments, I abandoned Hermitic conditions at the outset; it seemed illogical to retain them in an analysis which recognises sixteen different square roots of -1. We shall

determine the reality conditions of the present theory independently. By relativity considerations it is possible to determine them uniquely.

For a formal treatment, it is best to begin by considering the reality conditions for rotations. Corresponding to a matrix E_μ, we can consider two antithetic rotations, $E_\mu\theta_\mu$, $E_\mu iu_\mu$, where θ_μ and u_μ are real. Only one of these will be admissible for physically real phenomena. Now "physical reality" is an invariant property; therefore our reality condition must be such that it is invariant for all relativity rotations of the frame of reference, provided that these rotations are themselves physically real and therefore satisfy the reality condition that is being considered.

For trial suppose that a rotation $q=e^{\frac{1}{2}E_\mu\theta_\mu}$ is physically real, if $E_\mu\theta_\mu$ is imaginary. Then q is in general complex. Consider another rotation $q_1=e^{\frac{1}{2}E_\nu\theta_\nu}$, whose matrix $E_\nu\theta_\nu$ is imaginary and therefore satisfies the proposed reality condition. If now we apply the physically real rotation q to the frame of reference, $E_\nu\theta_\nu\rightarrow q\,(E_\nu\theta_\nu)\,q'$. If E_μ anticommutes with E_ν, $q\,(E_\nu\theta_\nu)\,q'$ is complex, so that q_1 no longer satisfies the reality condition. Thus the proposed reality condition is non-invariant, and must be rejected.

Accordingly the reality condition for a rotation is that q must be real. Then the matrix $E_\nu\theta_\nu$ of the rotation q_1 is real, and remains real when it is transformed to $qE_\nu\theta_\nu q'$ by the real rotation q.

The essential point in the argument is that the matrix of a relativity rotation can be transformed by applying another relativity rotation; so that any proposed reality condition is employed twice over and its self-consistency is thereby put to a test. In other words we have to secure that the physically real rotations constitute a Group.

Considering the most general rotation in four-dimensional space-time†

$$q=\exp\tfrac{1}{2}\,(E_{23}\theta_{23}+E_{31}\theta_{31}+E_{12}\theta_{12}+E_{14}iu_{14}+E_{24}iu_{24}+E_{34}iu_{34}),\quad(6\cdot12)$$

the condition that q is real requires that θ_{23}, θ_{31}, θ_{12}, u_{14}, u_{24}, u_{34} shall all be real. An E_{14} rotation gives, as in (4·15),

$$x_1'=x_1\cos(iu_{14})-x_4\sin(iu_{14}),\quad x_4'=x_1\sin(iu_{14})+x_4\cos(iu_{14}),$$

or, if $x_4=it$,

$$x_1'=x_1\cosh u_{14}+t\sinh u_{14},\quad t'=x_1\sinh u_{14}+t\cosh u_{14},\quad(6\cdot13)$$

so that the rotation is hyperbolic (Lorentz transformation). Thus the relativity transformations in four dimensions consist of three circular rotations θ and three Lorentz transformations u—in agreement with experience.

As a by-product we see that the real quantity concerned in the Lorentz transformations is $t=x_4/i$, so that x_4 is imaginary (if x_1, x_2, x_3 are real). We have thus a deductive proof of the result, already noticed, that (6·11) is a

† We here exhibit the six elementary rotations collected together for reference. As explained in § 4·2, it would be necessary to restrict them to infinitesimal rotations if it were intended to apply them simultaneously.

monothetic matrix. It is a matter of convention that it is imaginary, not real. By taking x_1, x_2, x_3 real, so that x_4 is imaginary, we conform to the convention of ordinary relativity theory which assigns real measure to time-like intervals. But we might equally have taken x_1, x_2, x_3 imaginary so that x_4 would be real; ds is then real for space-like intervals. Thus the compulsory reality conditions are:

$$\left.\begin{array}{l}\text{For rotations (in four dimensions): } q \text{ is a real matrix,}\\ \text{For intervals or displacements: } d\mathbf{s} \text{ is a monothetic matrix.}\end{array}\right\} \quad (6{\cdot}14)$$

Alternatively we may measure an interval by its vector density

$$d\mathfrak{s} = d\mathbf{s}\sqrt{-g} = d\mathbf{s}.iE_5.$$

Then by (6·11) $$-id\mathfrak{s} = E_{15}dx_1 + E_{25}dx_2 + E_{35}dx_3 + E_{45}dx_4. \quad (6{\cdot}15)$$

Since E_5 is real, $d\mathfrak{s}$ is antithetic to $d\mathbf{s}$; and with the usual convention ds is real and $d\mathbf{s}$ is imaginary.

At this stage it is well to review the progress of our theory of space and time. We have shown that, starting with a basal wave vector ψ, it is possible to construct a continuum of "equivalent" points which forms a four-dimensional hypersphere in five dimensions. In Chapter IV we showed that this continuum has the local isotropic quality of ordinary space-time in that rotations in any of its coordinate planes are relativity rotations; but we did not there discriminate between circular and hyperbolic rotations. We have now confirmed the resemblance in greater detail by showing that the relativity transformations of this theoretical continuum consist of three ordinary rotations and three Lorentz transformations; or equivalently that one of the four dimensions is antithetic to the other three.

Further, referred to this continuum, the basal vector ψ has the transformation properties of Dirac's ψ, and in fact satisfies identically an equation identifiable with Dirac's wave equation.

We must remind ourselves, however, that this is no more than the embryo of the actual macroscopic space-time of our experience. In the next section we shall find a very significant difference which shows the need for introducing further developments in due course.

6·2. Translations.

Considering the neighbourhood of a particular point P on the hypersphere, we take as usual the coordinate x_5 to be along the radius at P, so that x_1, x_2, x_3, x_4 are rectangular coordinates in space-time. This coordinate system is necessarily local, for it is impossible to construct an extended system of rectangular coordinates in a curved space. We must therefore restrict ourselves to an infinitesimal region around P.

The transformation $q = e^{\frac{1}{2}E_{15}\theta}$ gives a rotation of the hypersphere in the plane x_1x_5 and therefore displaces P in the x_1 direction. Thus infinitesimal

translations of a point along the four axes in space-time correspond to relativity rotations with matrices E_{15}, E_{25}, E_{35}, E_{45}.

The recognised relativity transformations in space-time, viz. four translations, three rotations in space and three Lorentz transformations, make up the ten relativity rotations in five dimensions introduced in § 4·3. The customary approximation which treats space-time as flat hides the fact that translation is a form of rotation, viz. rotation about the centre of curvature of space-time.

The general translation in space-time is accordingly given by the transformation

$$q = \exp \tfrac{1}{2}(E_{15}\theta_{15} + E_{25}\theta_{25} + E_{35}\theta_{35} + E_{45}\theta_{45}). \tag{6·21}$$

If we apply the previous reality condition that q is real, the expression in real variables is

$$q = \exp \tfrac{1}{2}(E_{15}iu_{15} + E_{25}iu_{25} + E_{35}iu_{35} + E_{45}\theta_{45}), \tag{6·22}$$

since E_{15}, E_{25}, E_{35} are imaginary, and E_{45} is real. As in (6·13) the u's give hyperbolic rotations, and θ_{45} gives a circular rotation. This means that the continuum is open in three (space) dimensions and closed in one (time) dimension—the reverse of the conditions in actual space-time.

The root of the trouble is that, when space-time is pictured in five dimensions, these dimensions are four space-like and one time-like; for the radius of curvature represented in the fifth dimension is space-like. But the matrices associated with these dimensions are three imaginary and two real. The real matrix E_5 is incongruously associated with a space-like dimension.

Some writers on relativity have mooted the possibility that the world might have negative curvature. (Negative curvature refers to the Gaussian curvature which is proportional to $1/R^2$. If R^2 is negative, R is an imaginary length, or equivalently it is a time-like radius.) But the proposal has been treated from the point of view of formal mathematics, and can scarcely be entertained in physical theory.†

We shall find the significance of this incongruity in the next section. But we shall first consider how it is to be reconciled with the invariance of reality conditions. To conform to the actual universe it will be necessary to admit that the exponent of q is real in (6·12) but imaginary in (6·21); that is to say, the general rotation Θ is to be resolved into $\Theta = \Theta_1 + \Theta_2$, where

$$\left.\begin{aligned} \Theta_1 &= E_{23}\theta_{23} + E_{31}\theta_{31} + E_{12}\theta_{12} + E_{14}iu_{14} + E_{24}iu_{24} + E_{34}iu_{34}, \\ \Theta_2 &= E_{15}\theta_{15} + E_{25}\theta_{25} + E_{35}\theta_{35} + E_{45}iu_{45}, \end{aligned}\right\} \tag{6·23}$$

with the reality condition that Θ_1 is real and Θ_2 is imaginary.

What then becomes of the invariance of physical reality? The saving circumstance is that we have restricted ourselves to an infinitesimal region

† The objection to unclosed space arises from quantum theory rather than from relativity theory. It will become obvious as our investigation proceeds.

round a point P; and at the point P the direction denoted by E_5 (normal to space-time) is *absolutely* distinguished from the other directions (§ 4·4). Thus the separation of Θ into Θ_1 and Θ_2 is defined in an absolute way; and we can attach different reality conditions to the two parts without coming into conflict with relativity principles.

Let us now apply a rotation $\Theta_1' + \Theta_2'$ to a rotation $\Theta_1 + \Theta_2$. Under the transformation $q_1 = e^{\frac{1}{2}\Theta_1'}$, Θ_1 and Θ_2 transform separately. By the above reality conditions q_1 is real, and therefore $q_1\Theta_1 q_1'$, $q_1\Theta_2 q_1'$ remain real and imaginary respectively, and continue to satisfy the reality conditions. Under the transformation $q_2 = e^{\frac{1}{2}\Theta_2'}$, Θ_1 and Θ_2 are not kept separate. The real matrix Θ_1 is transformed into a complex matrix $q_2\Theta_1 q_2'$; but it is easily verified that the imaginary terms in it are of the form Θ_2 so that they satisfy the reality conditions. Similarly the imaginary matrix Θ_2 becomes a complex matrix $q_2\Theta_2 q_2'$; but the real terms introduced are of the form Θ_1, and satisfy the reality condition for Θ_1. Thus the reality conditions are found to be self-consistent, and the rotations which satisfy them form a Group.

We have confined ourselves to an infinitesimal (initial) region because it is only locally that we can pick out a unique direction, absolutely different from other directions, to distinguish as E_5. But is it sufficient to treat an infinitesimal region? A relativity rotation cannot be real at some points and unreal at other points of space-time; we ought therefore to show that the same rotation tested in two different localities fulfils the proposed reality conditions at both or neither. It turns out that this self-consistency is assured automatically. If we examine the reality condition at another point of space-time, we must first make sure that it is a real point; and the test of its reality is that it is equivalent to the real point first considered—that it can be transformed into it by a relativity rotation which satisfies the reality condition. Thus we have to lay down the reality condition for relativity transformations of a single initial point (given as real) *before* we can decide what values of the coordinates represent real points, i.e. points which have real equivalence to a point known to be real. Thus we have not to show that our adopted reality condition is self-consistent over a predetermined real domain; the domain over which it is self-consistent is *ipso facto* the domain of real points.

We begin with one real point P—the observer, in fact, for all reality is relative to him. We determine a group of infinitesimal transformations which we define as the physically real transformations. These transformations applied to P give all the neighbouring real points. By a process of continuation (using at each stage the local reality conditions) we reach the more distant real points. Any of these real points can now be taken as the initial real point; proceeding from it we shall by the same construction obtain the

same real domain. This is a consequence of the group property of the transformations used.

We have shown that the reality conditions (6·23) satisfy relativity requirements; but we have still to explain why these are the conditions prevailing in nature, rather than the simple condition that Θ_1 and Θ_2 should both be real.

6·3. Neutral Space-time.

We shall find later that the entity represented by a simple wave tensor is an electric particle (proton or electron). Up to the present we have dealt with one particle only; we have not yet developed the apparatus of description for two or more particles. It is therefore rather premature to talk about space-time, which is a macroscopic conception presupposing vast numbers of particles. What we have investigated is a preliminary geometrical framework in which the characteristics of a single elementary particle are represented vectorially. This framework is the genesis of macroscopic space-time; and it already contains two of the most essential features: (1) four dimensions, of which one is antithetic to the other three, and (2) a radius of curvature.

For clearness we shall here anticipate some of the changes which will take place in converting this preliminary conception into the space-time of macroscopic experience. By introducing great numbers of particles the radius of curvature will be greatly increased, relatively to the linear scale characteristic of a single particle (commonly recognised as the wave length of its Schrödinger waves). The increase of the population will give scope for irregularity, and the hypersphere of space-time will be distorted by gravitational fields. But the difference that chiefly concerns us now is that there will be a balancing of positive and negative particles. A universe containing only one particle, and therefore only one sign of charge, is lop-sided compared with a universe containing equal numbers of positive and negative charges. It will readily be imagined that if matter consisted of electrons only, the enormous negative potential would so alter the world that we should require a different type of space-time to frame the phenomena. It is just such a world in miniature that our theory of a single charged particle imitates.

The Riemannian space-time of Einstein's general relativity theory is derived from the extensional relations of *neutral* matter. The test bodies whose behaviour determines its characteristics—scales, clocks, moving particles, light waves—are electrically neutral. Moreover, all macroscopic matter in our experience is to a very high approximation neutral; if the proportion of electrons to protons differs from equality by one in a billion, the electric charge expressed in ordinary units is stupendous.

The test bodies are used to measure intervals, and hence determine the tensor $g_{\mu\nu}$ characteristic of neutral space-time. Ideally their use is restricted to regions where there is no electromagnetic field. In practice this is not a very important restriction, because the strongest electromagnetic fields encountered in nature correspond to a very trifling lack of balance of distribution of positive and negative charges. But in an electromagnetic field too strong to be neglected, it would be impossible to use the test bodies to determine $g_{\mu\nu}$, because there is no agreed definition as to how $g_{\mu\nu}$ is related to the indications of the test-bodies when electromagnetic fields are present. Each investigator has defined it according to his own fancy. No experimenter would undertake to make accurate measurements of length in an intense electromagnetic field; statements about lengths and distances within the field are inferences from observations made outside the field, and depend on the theoretical formulae employed in calculating the inference. Current scientific literature abounds in rival formulae (usually embodied in an "action-principle") for making such calculations, each corresponding to a different definition of $g_{\mu\nu}$ in regions where there is no means of determining it by direct observation.†

We must therefore consider even at this early stage the main difference between the space-time of a universe whose content is neutral and the space-time of a universe containing a particle or particles of one sign only. We shall call the space-time of a universe containing only positive particles *positively saturated*; if containing negative particles only, it is *negatively saturated*.

In § 3·9 we have shown that there exist two kinds of frame, right- and left-handed, which cannot be transformed into one another by a relativity rotation. If we take two vectors $T=\Sigma t_\mu E_\mu$, $T'=\Sigma t_\mu F_\mu$, where E_μ, F_μ are respectively right- and left-handed, these cannot be transformed into one another by a relativity rotation. Clearly a distinction of this kind is required to discriminate between positive and negative charged particles. We shall therefore provisionally identify T and T' with the complete stream vectors of positive and negative particles—subject, of course, to confirmation by detailed examination of the resulting properties.

Take right- and left-handed frames related as in (3·92), so that

$$F_{16}, F_5, F_{15}, F_{25}, F_{35}, F_{45} = -E_{16}, -E_5, -E_{15}, -E_{25}, -E_{35}, -E_{45}. \tag{6·31}$$

For the second particle we have $T'=\Sigma t_\mu F_\mu=\Sigma t_\mu' E_\mu$, where

$$t_{16}', t_5', t_{15}', t_{25}', t_{35}', t_{45}' = -t_{16}, -t_5, -t_{15}, -t_{25}, -t_{35}, -t_{45}, \tag{6·32}$$

the other ten components being the same for both particles. Hence, considering the position vector, the coordinates of the two particles‡ are

† See § 13·4.

‡ More strictly coordinates of elements of their probability distributions (§ 5·8).

respectively $(t_1, t_2, t_3, t_4, t_5)$ and $(t_1, t_2, t_3, t_4, -t_5)$. If we take as usual the E_5 axis normal to the small region that we are considering, so that t_1, t_2, t_3, t_4 are infinitesimal, the particles are in antipodal regions in five dimensions; but the ordinary four-dimensional point of view is that they are at the same point of space-time (t_1, t_2, t_3, t_4),† and that the centre of curvature of space-time is in opposite directions along the normal, according to which particle we are considering.

There is nothing surprising in this. The two particles are contemplated as *alternatives*; for our analytical machinery is not yet capable of dealing with two particles at once. By Einstein's theory the curvature of space-time depends on its contents. So, if we start by considering an infinitesimal region, we do not know how the region will continue until we have decided on the contents of the region. It will curve away from the tangent plane more or less strongly according as the density is high or low. We now see further that it will curve to one side of the plane or the other side of the plane according as we insert an elementary positive or negative charge (or a probability thereof). This is not noticed in Einstein's theory, because that is a theory of macroscopic matter, which (even if it is electrically charged) contains equal numbers of positive and negative charges to an extremely high approximation.

It will be seen that the idea first suggested by the five-dimensional picture —that the positive and negative particles are at antipodal points on a fixed sphere—is rather misleading. They are not simultaneously present; and the sphere is not fixed until we have decided *which* is present.

Our result is that, considering an infinitesimal region of space-time, if it contains part of the probability distribution of a positively charged particle, the radius of curvature will be in one direction of the normal, say x_5; and if it contains part of the probability distribution of a negatively charged particle, the radius of curvature will be in the opposite direction $-x_5$; or briefly, if the region is positively saturated the radius of curvature will be in the direction x_5, and if negatively saturated it will be in the direction $-x_5$. Suppose now that it is neutral—having equal probability of positive or negative charge. Two possibilities are open. The curvature may be zero, or the radius of curvature may be in the direction of the imaginary normal ix_5. These are the only alternatives which have neither a positive nor a negative bias. We know from ordinary relativity theory that the first alternative is incorrect; neutral matter *does* involve a curvature of space-time. Therefore we must accept the second alternative; the radius of curvature is in the direction ix_5. That is to say the radius of curvature changes its character from time-like to space-like when we pass from a positively or negatively saturated world to a neutral world.

† We have chosen the relation (3·92) rather than (3·93) or (3·94) in order that our formulae may refer to particles at the same point of space-time.

This result will be confirmed in our subsequent investigations. But I think it is clear, even at this early stage, that it is forced upon us.

We can now understand the origin of the curious reality conditions in § 6·2. The whole trouble was that the radius of curvature was associated with a real matrix E_5 indicating time-like character. But the radius of curvature *is* time-like in the positively or negatively saturated world to which the most elementary formulae relate. It is in the transition to a neutral world that the space-like radius of curvature becomes substituted for a time-like radius. In a positively saturated world the relativity transformations $q = e^{\frac{1}{2}\Theta}$ satisfy the simple condition that Θ is a real matrix. The $\sqrt{-1}$ never gets a footing in so simple a world. According to this theory a positively saturated world is open in its space dimensions and closed in its time dimension. Saturation is so remote from the conditions of our actual experience that we certainly cannot bring forward any observational evidence to the contrary. We can well imagine that the stupendous electrical repulsions would be sufficient to burst any closed space.

Since the vector $E_5 x_5$ towards the centre of curvature has in neutral space-time a value antithetic to its value in electrically saturated space-time, it is antithetic to the four-dimensional position vector

$$X = E_1 x_1 + E_2 x_2 + E_3 x_3 + E_4 x_4 .$$

The rotation about a centre at $E_5 x_5$, which produces a displacement dX, is given by $dX/E_5 x_5$. Thus the rotation Θ_2 is antithetic to its value in electrically saturated space-time, and is therefore imaginary. This gives the reality conditions (6·23). The change from a time-like to a space-like radius of curvature—from saturated to neutral space-time—is the source of the $\sqrt{-1}$ which is such an inescapable feature of quantum formulae.

6·4. Kinematical and Electrical Matrices.

For general developments it is more convenient to take right- and left-handed sets E_μ, F_μ connected by (3·93),† so that

$$F_{0\mu}, F_{16} = -E_{0\mu}, -E_{16} \quad (\mu = 1, 2, 3, 4, 5), \tag{6·411}$$

the other ten matrices being the same. Let

$$N_{0\mu}, N_{16} = iE_{0\mu}, iE_{16}. \tag{6·412}$$

We shall call N_μ a *neutral set*. As a further generalisation we define a *macroscopic set* M_μ by

$$M_{0\mu}, M_{16} = \lambda E_{0\mu}, \lambda E_{16}, \tag{6·413}$$

the other ten matrices being the same. Then λ is a scale constant which may be real, imaginary or complex. Real values correspond to electrically saturated space-time, and imaginary values to neutral space-time. An

† We shall see in § 6·5, that (3·92) gives the association of vector densities and (3·93) the association of vectors.

ordinary field of positive or negative potential could therefore be represented by a complex value of λ, the real part being very small compared with the imaginary part; this kind of representation is only used when we are pursuing a unified field theory, involving non-Riemannian geometry. More usually macroscopic electrical fields are represented as perturbing influences superposed on Einstein's neutral space-time.

We may express a complete space vector in terms of macroscopic components, namely

$$T = \Sigma m_\mu M_\mu = \Sigma_k m_\mu E_\mu + \lambda \Sigma_e m_\mu E_\mu, \tag{6·42}$$

where Σ_k refers to the ten unchanged matrices and Σ_e to the six changed matrices in (6·413). We shall call the unchanged matrices the *kinematical matrices*, and the matrices whose sign is changed between right- and left-handed frames the *electrical matrices*. The electrical matrices are accordingly

$$E_1, E_2, E_3, E_4, E_5, E_{16} \tag{6·43}$$

and the kinematical matrices are those associated with the ten rotations in five dimensions. We shall adhere to this nomenclature, irrespective of the physical interpretation.†

The relativity rotations in five dimensions correspond to the ten kinematical matrices. We may therefore distinguish them as *kinematical rotations*, the other six relativity transformations being *electrical rotations*. As shown in § 4·3 the kinematical and electrical parts of (6·42) are transformed separately by the kinematical rotations; so that from the ordinary standpoint they are separate vectors arbitrarily combined into one analytical expression. In combining them we can introduce a scale constant λ, which remains invariant in the transformations. For example, in macroscopic spherical space-time we may meet with two distinct but somewhat analogous vectors, say a velocity and a spin. We think it probable that there is some significant combination of these into a complete space vector, which will lead to a more far-reaching theory of the phenomena. But until the details of that theory are worked out, the combination involves an undetermined scale ratio λ. If we regard the components m_μ of the complete vector as definite and equal to the components of the two separate vectors as ordinarily measured, λ must be embodied in the frame of our ordinary measurements which then becomes a macroscopic frame M_μ.

We have found representations of a complete space vector in five dimensions (§ 4·3) and in four dimensions (§ 4·4). There is also a representation in six dimensions. The components $t_1, t_2, \ldots t_{15}$ form a 15-vector, or antisymmetrical tensor of the second rank in six dimensions. From a purely algebraic standpoint this is the most fundamental representation; and the

† The matrix associated with a particular characteristic depends on whether the characteristic is expressed as a vector, vector density or strain vector. Thus the "electrical matrix" E_1 has not always an electrical significance.

theory of E-symbols is primarily a study of the group of rotations in space of six dimensions. But the ordinary vectors of that space (6-vectors) cannot be expressed in terms of E-symbols. The physical application of this group is specialised by the fact that space, as defined in physics, is an abstraction of the extensional relations of *neutral* matter, and therefore involves the superposition of a right- and a left-handed frame in the 6-space. This superposition involves the selection of a pentad (here taken to be $E_{0\mu}$) which, together with E_{16}, is reversed in sign between right- and left-handed frames. One suffix (in this case 0) thereby acquires distinctive properties; and, once chosen, it remains associated with an invariant direction in the 6-space. The remaining five dimensions become the 5-space of § 4·3.

To review the argument:—A right-handed frame can be represented as having complete symmetry in a 6-space. Such a frame would be appropriate as a reference system if the universe consisted of particles of one sign only. Since actual physical systems consist of nearly equal numbers of positive and negative particles, our actual reference system is based on a frame which is a superposition of right- and left-handed frames, the one being, as it were, a reflection of the other. In the composite frame it is no longer true that all directions in six dimensions are equivalent; one direction must be chosen as the axis of the reflection, and thereby becomes distinguished from the others. We have associated the suffix 0 with this direction. Thus, whilst the elementary right-handed frame exhibits six-dimensional relativity, the actual composite frame exhibits five-dimensional relativity. Accordingly the starting point of the theory of actual space-time is the five-dimensional representation of wave tensors treated in § 4·3.

The kinematical rotations have the same relations to the frames E_μ, F_μ, N_μ, M_μ. In each case they rotate separately the two parts into which (6·42) is divided, and the value of λ does not affect the result. But the electrical rotations have opposite effects on E_μ and F_μ, and therefore correspond to some kind of separation of electric charge, or polarisation. For this reason they give apparently non-relativistic transformations of neutral or macroscopic space vectors. This does not mean that it is unprofitable to investigate them further; but it is only by dropping the usual representation in Riemannian space-time, and following a unified geometrical representation of gravitational and electromagnetic fields, such as those of Weyl and the author, that these electrical transformations become admissible.

6·5. Summary of the Reality Conditions.

Since physical quantities can be expressed as space vectors, vector densities (four-dimensional), or strain vectors (three-dimensional densities), we have to state the reality conditions separately for these three forms. The following conditions refer to vectors, etc., in neutral space-time.

(1) *Space vectors.*

The condition is that the electrical part, involving the matrices

$$E_{16}, E_1, E_2, E_3, E_4, E_5,$$

is antithetic to the kinematical part; or T_e is antithetic to T_k.

(2) *Vector densities.*

Since the vector density $\mathfrak{T} = iTE_5$, the above condition becomes

$$(\mathfrak{T}E_5)_e \text{ is antithetic to } (\mathfrak{T}E_5)_k,$$

so that the part of $\mathfrak{T}$, involving the matrices

$$E_{16}, E_{15}, E_{25}, E_{35}, E_{45}, E_5,$$

is antithetic to the rest of the expression.

(3) *Strain vectors.*

Since the strain vector or three-dimensional density $S = iTE_{45}$, the condition becomes $(SE_{45})_e$ is antithetic to $(SE_{45})_k$.

By (4·66) the part of S which corresponds to the electrical part of T is that containing

$$E_{23}, E_{31}, E_{12}, E_4, E_5, E_{45},$$

i.e. the real matrices. Thus the part of S containing real matrices is antithetic to the part containing imaginary matrices. We have therefore the simple result that the *coefficients* s_μ of a strain vector are homothetic.

(4) *Kinematical rotations.*

These are restricted to ten components and their reality conditions, given in full in (6·23), secure that the part containing

$$E_{15}, E_{25}, E_{35}, E_{45}$$

is antithetic to the rest. Comparing with the above results, we see that *the matrix of a rotation must be regarded as a vector density.*

These conditions are founded on the result obtained in § 6·3, that for a position vector $E_5 x_5$ is antithetic to $E_1 x_1 + E_2 x_2 + E_3 x_3 + E_4 x_4$. By § 5·8, a position vector is part of a complete *vector density.* We have therefore to choose right- and left-handed frames related in such a way that the reality condition (2) for a vector density agrees with this. This justifies the choice made at the beginning of § 6·4 when we took the connection to be that given by (3·93).

The foregoing are the natural reality conditions in neutral space. But it must be understood that, since the only test is the invariance of physical reality, the conditions become less stringent as we limit the variety of transformations contemplated. Thus if we have a 5-vector U and a 10-vector V, forming a complete space vector $U + V$ which satisfies the foregoing reality conditions, the combination $U + iV$ will violate them. But so long as

we confine ourselves to the ten kinematical rotations which transform U and V separately, there is no special reason for preferring the combination $U+V$ to $U+iV$. It may well happen that $U+iV$ is the customary form of combination. All we can say is that this latitude exists, because the expression is a mental association rather than a genuine combination. Combination implies some loss of independence of U and V; so that a more general transformation is contemplated in which the combination $U+V$ persists and the combination $U+iV$ is broken up.

The hamiltonian $H=E_1p_1+E_2p_2+E_3p_3+E_4p_4-m$ adopted in Chapter v satisfies the reality conditions for a vector density, since $-m$ is the component $E_{16}im$. Accordingly $J=\psi\chi^*$ is a vector density. If we require the wave equation for a vector we must identify (p_1, p_2, p_3, p_4) with the adjoint vector $(j_{15}, j_{25}, j_{35}, j_{45})$, and therefore write the hamiltonian as

$$H=E_{15}p_1+E_{25}p_2+E_{35}p_3+E_{45}p_4-m. \tag{6·51}$$

This satisfies the conditions (1) for a space vector. We need not stop to decide which of these forms is the most advantageous, because we shall find later that the form chiefly required in practice is that which corresponds to a strain vector.

6·6. Charge and Spin.

We have found in (5·83) that the complete stream vector of an elementary particle is, if the axes are suitably chosen, of the form

$$J_a=(E_1+E_{23}+E_{45}+E_{16})\,im. \tag{6·61}$$

We have set $\alpha=im$ in anticipation of the identification found below. We first notice that, since E_{23}, E_{45} are real and E_1, E_{16} are imaginary, J_a satisfies the reality conditions for a space vector in neutral space-time (§ 6·5 (1)).

It may seem surprising that the result obtained in § 5·8 is already adapted to neutral space-time, which was not formally introduced until later. The reason is that in the course of the derivation we selected the combinations of suffixes *in accordance with experience*; and since we have no experience of electrically saturated space-time it was implicitly excluded. For a pure wave tensor in electrically saturated space-time the appropriate form is

$$(E_1+E_{24}+E_{35}+E_{16})\,im,$$

which is monothetic. It is easily seen that the momentum vector is then in the space-like direction E_3.

The hamiltonian is given by the last two terms of (6·61):

$$H_a=E_{45}im-E_{16}im=E_{45}ip_0+m. \tag{6·62}$$

We thus verify that m is the proper mass. The terms E_{45} and E_{16} in (6·61) accordingly represent the energy and the proper energy, which, although

equal in magnitude (owing to our special choice of coordinates), are exhibited separately in the wave tensor.

To obtain a particle of opposite sign to J_a we must reverse the sign of the electrical matrices, obtaining

$$J_b = (-E_1 + E_{23} + E_{45} - E_{16})\, im.$$

The corresponding hamiltonian is $H_b = E_{45} ip_0 - m$. It will be seen that if the standard form of the hamiltonian is taken to be

$$E_{15}p_1 + E_{25}p_2 + E_{35}p_3 + E_{45}ip_0 - m,$$

the identification of m with the proper mass must be qualified by the proviso that this refers to the magnitude, and that the sign of m is that of the charge. The identification was based on (5·24) which gives m^2, not m.

The terms E_1 and E_{23} describe a spin, the one giving the axis of the spin and the other the plane of the spin. But they play somewhat different parts, because one is reversed when we change the sign of the charge and the other is not. Clearly E_{23}, which is not reversed, represents the mechanical spin; and E_1, which is reversed, represents the magnetic moment.

To obtain a particle of opposite spin we reverse the sign of E_{23}. In order that the wave tensor may remain pure, we must also reverse the sign of another term (§ 5·5). We cannot reverse E_{16}, since that would reverse the charge; we cannot reverse E_{45}, since that would give negative energy p_0, representing a "minus-particle". Hence we must reverse E_1. This confirms our interpretation of E_1 as a magnetic moment, whose sign depends on the direction of spin.

Assuming arbitrarily that J_a represents a positive charge with a "positive" direction of spin in the plane x_2x_3, we have the following classification:

$$\begin{aligned}
J_a &= im \quad (E_1 + E_{23} + E_{45} + E_{16}) && \text{positive charge, positive spin,}\\
J_b &= im\,(-E_1 + E_{23} + E_{45} - E_{16}) && \text{negative charge, positive spin,}\\
J_c &= im\,(-E_1 - E_{23} + E_{45} + E_{16}) && \text{positive charge, negative spin,}\\
J_d &= im \quad (E_1 - E_{23} + E_{45} - E_{16}) && \text{negative charge, negative spin.}
\end{aligned} \tag{6·63}$$

The three-dimensional densities or strain vectors are also of importance. They are given by $S = iJE_{45}$. We obtain

$$\begin{aligned}
S_a &= -im \quad (E_1 + E_{23} + E_{45} + E_{16}) && \text{positive charge, positive spin,}\\
S_b &= -im \quad (E_1 - E_{23} - E_{45} + E_{16}) && \text{negative charge, positive spin,}\\
S_c &= -im\,(-E_1 - E_{23} + E_{45} + E_{16}) && \text{positive charge, negative spin,}\\
S_d &= -im\,(-E_1 + E_{23} - E_{45} + E_{16}) && \text{negative charge, negative spin.}
\end{aligned} \tag{6·64}$$

We notice that in the strain vector the mechanical spin is represented by E_1 and the magnetic moment by E_{23}—opposite to the representation in the space vector.

Since
$$S_a+S_b+S_c+S_d=4m, \tag{6·65}$$
the four forms can be regarded as resulting from a spectral analysis (§ 5·7) of algebraic numbers. The properties of matter which is neutral as regards charge and spin can be represented by simple algebraic wave functions of the type introduced by Schrödinger. To introduce charge and spin Schrödinger's algebraic wave tensors are analysed into components S_a, S_b, S_c, S_d, which can then be assigned modified probabilities independently, so that they no longer balance. This is the simplest way of connecting the theory of protons and electrons with the ordinary relativistic mechanics of neutral matter, as will be shown in § 13·6.

An entity represented by $\frac{1}{4}(J_a+J_b+J_c+J_d)$ or by $\frac{1}{4}(S_a+S_b+S_c+S_d)$ will be called a *neutral particle*. The term is not intended to have any connection with the neutron. A neutral particle is not a combination of four particles; it is a single particle which (so far as our information goes) has equal probability of being a proton or electron and equal probability of either direction of spin. We regard S_a, S_b, S_c, S_d as four elementary states of a particle. In general the probabilities of the four states will be different. As a particular case (which, however, is the commonest case in practice) the probability of one of the states, say S_a, may be unity; the particle is then classed definitely as a proton of positive spin.

From the ordinary standpoint a neutral particle is a mathematical fiction, having no counterpart in experimental physics. Usually, if an experimenter knows anything at all about a particle, he knows whether it is a proton or electron. Direction of spin is less easily recognised; and a combination S_a+S_c, which gives a particle neutral as regards spin, may often represent the experimental knowledge available.

The complete stream vector (space vector) of a neutral particle consists of a single component $E_{45}t_{45}$, and its three-dimensional density (strain vector) consists of a quarterspur $E_{16}s_{16}$. Thus we have a very simple way of passing from electrical to neutral particles, namely by taking the quarterspur of the strain vector—an operation which corresponds to *contraction* of the corresponding wave tensor.

In more general coordinates the term E_{23} in the strain vector is replaced by three components E_{23}, E_{31}, E_{12}. We have seen that these represent the magnetic moment. The electric moment (if any) will be represented by terms E_{14}, E_{24}, E_{34}, since by the usual electromagnetic equations the magnetic and electric moments form a 6-vector. No such terms occur if the particle is at rest in the coordinate system; this is what we should expect,

since an electric moment implies a doublet, and cannot be associated with a single particle at rest. Electric moment terms are introduced if we set the particle in motion by applying a Lorentz transformation with matrix E_{24} or E_{34} to (6·63). The electric moment thus produced is real. Dirac's theory seems to differ from ours on this point, since he obtains an *imaginary* electric moment for the electron.†

6·7. Minus-particles.

The wave tensor $-J_a$ must presumably be regarded as expressing the absence of a particle of the kind represented by J_a, or else an entity which is observationally equivalent to such absence. We may accept the current view (due to Dirac) that positrons and negatrons are minus-particles of this kind equivalent to the absence of electrons and protons. Thus

$$-J_a = im(-E_1 - E_{23} - E_{45} - E_{16})$$

represents a negatron. The sign of the last term shows that the charge is negative.

The sign of the third term should signify that the energy is negative, according to the somewhat confusing definition of energy in wave mechanics. But the energy of a negatron or positron, as ordinarily understood, is certainly positive. It is necessary to clear up this discrepancy of definition.

Energy, momentum and spin are familiar conceptions in classical mechanics, and a complete energy-momentum-spin vector can be defined for a macroscopic system. Let us suppose that, by adapting this definition, we can assign to an elementary particle a complete vector T representing its energy, momentum and spin "as ordinarily understood". In wave mechanics we associate with the elementary particle a wave tensor $J_a = \psi\chi^*$, whose components have at least some analogy with the energy, momentum and spin of classical mechanics. It is tempting to assume that $T = J_a$. But J_a has the idempotent property that, apart from a numerical factor depending on the choice of units, $J_a^2 = J_a$. We have therefore just as good reason to make the identification $T = J_a^2$.

The proton and negatron which have opposite stream vectors J_a, $-J_a$, have the same vector J_a^2. The identification $T = J_a^2$ will accordingly make the energy of negatrons (and positrons) positive, as it should be.

We therefore regard the primary wave tensor $J = \psi\chi^*$ of an elementary particle as a charge-current vector, and J^2 as the true energy-momentum vector. We have been considering the vector J for simplicity, but strictly the relation is between the three-dimensional densities S, S^2; the charge-current density is represented by S, and the energy-momentum density by S^2.

† *Quantum Mechanics*, 2nd ed., p. 263.

Since S^2 is an algebraic multiple of S, the two vectors coalesce for an elementary particle. This has happened by design rather than by accident. The coalescence of S^2 with S is virtually the definition of an elementary particle; for elementary character has been taken to correspond to purity, and purity to idempotency. This helps us to understand why so much importance is attached in wave mechanics to resolution into pure or idempotent constituents. It is for these constituents that we are able to replace a quadratic dynamical property by a linear property. For impure constituents these would be distinct wave tensors describing different properties. Broadly speaking S^2 represents mechanical characteristics and S electrical characteristics; they become unified only in an elementary particle whose momentum and current coincide.

This agrees with general relativity, in which the mechanical properties are specified by an energy tensor of the second rank (a quadratic function of the individual velocities), and the electrical properties by a charge-current vector (a linear function of the individual velocities).

CHAPTER VII

STRAIN VECTORS AND PHASE SPACE

7·1. Internal Wave Functions.

In classical mechanics it is usual to resolve the motion of a system of particles into a motion of the centre of mass, and a motion of the individual particles relative to the centre of mass. We shall distinguish these as the *external* and the *internal* motions of the system.

Similarly in wave mechanics we resolve the motion of a system into an external motion specified by external wave functions, and an internal motion specified by internal wave functions. So far as the external motion is concerned, the system is equivalent to a single particle located at the centre of mass. It is characterised dynamically by an external momentum vector (p_1, p_2, p_3, p_0) and a proper mass m. If the space-time frame of reference for the external motion is changed, the momentum vector undergoes rotations and Lorentz transformations. The theory of the external wave function coincides with that of a simple particle.

The internal wave function introduces new ideas. Lorentz transformations are not applicable to the internal motion; for, by definition, the internal motion is relative to the centre of mass, and, if we applied a Lorentz transformation to it, it would be referred to some other standard of rest. *The time-axis of the frame of reference for internal motions and wave functions must agree with the direction of the external momentum vector.*

Thus we have a uniquely defined space-time frame for internal motions, and "simultaneity" has a definite meaning—provided that the system is not so extensive as to make it necessary to take account of curvature of space-time. It is part of our mental conception of a complex system that it is a simultaneous aspect of its several parts. Each particle has three internal (relative) coordinates $\xi_\mu = x_\mu - \bar{x}_\mu$; and the momenta conjugate to ξ_μ are the internal momenta of the particles of the system. There is one time-coordinate s common to the whole system. If the same coordinate system is used for the internal and external motions, the external time will also be s. But in general different frames will be used, since it would be idle to consider external motion if the external frame had always to be chosen so that the system was at rest in it; the external time t will then differ from the proper time s.

The independent variable for the internal motions and wave functions is always the proper time s. It should be noticed that if, for special purposes, separate internal time-coordinates analogous to the internal space coordinates are assigned to the particles, these will be $\tau = t - \bar{t}$. To associate t

with $x_\mu - \bar{x}_\mu$ (as is sometimes attempted) is a hybrid procedure unwarranted by any theoretical principle.

For internal wave functions Lorentz transformations are definitely ruled out; but relativity rotations in three-dimensional space are applicable. The direction of the time-axis is prescribed by the external momentum vector; but there is no corresponding specification of the orientation of the other axes.

The use of internal wave functions which, by their very nature, cannot be subjected to Lorentz transformations is often called "non-relativistic treatment"—with the implication that it conflicts with the principle of relativity. This is a misunderstanding of the nature of the Lorentz transformation and its place in relativity theory. So long as we deal with quantities defined as relations between physical entities—relative coordinates, relative velocities, relative momenta—we are on safe ground. (By relative coordinates we do not, of course, mean the difference of coordinates in some arbitrary frame of reference,† but the coordinates of one particle in the frame in which the other particle is at rest at the origin.) It is when we introduce into our formulae analogous quantities relative, not to physical reference objects, but to abstract frames of space and time, that Lorentz invariance is demanded. For we are then unable to define which of the equivalent frames has been selected—to which frame our formulae apply—so that, if the formulae are to mean anything at all, they must have a form invariant for all transformations of the frames.

In relativity theory itself there has been no such tendency to let Lorentz invariance grow into an obsession. One of the best known formulae in relativity theory is $ds^2 = -\gamma^{-1}dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta\, d\phi^2 + \gamma dt^2$ for the line element in the gravitational field of a particle. This is not invariant for Lorentz transformations; but we can scarcely describe the formula which is the source of the three crucial tests of Einstein's theory as "non-relativistic". If Lorentz invariance is not demanded in the investigation of the motion of a planet round the sun, it can scarcely be demanded in the investigation of the motion of an electron round a nucleus.

Hitherto, in developing the analytical theory, we have had in mind the motion of a particle or the external motion of a system. In the present chapter we shall introduce the modifications appropriate for treating the internal configuration and motion of a system. There are two fundamental differences. Firstly, the time reckoning will now be a proper time, fixed by the external momentum vector, and therefore invariant for any permissible transformations of the variables describing the internal configuration. Secondly, whereas change of the coordinates $\bar{x}_\mu$ of the centre of mass is a

† This possible confusion of meaning does not arise in the case of relative momentum, which obviously has no connection with difference of momentum.

relativistic change, all points in the space-time hypersphere being equivalent, change of the relative coordinates ξ_μ is in general a strain or intrinsic distortion of the system, and is therefore not to be represented as a relativity rotation of the vectors describing the internal state.

7·2. Covariant Wave Tensors.

We shall now consider the transformations of a covariant wave tensor of the second rank $S_{\alpha\beta}$. Like the mixed tensor it is a matrix composed of 16 elements; and it may be resolved into matrix components in an orthogonal frame by the same formula $S=\Sigma s_\mu E_\mu$; but s_μ will not be a space vector. We shall call s_μ a *strain vector*. Thus a strain vector corresponds to a covariant wave tensor in the same way that a complete space vector corresponds to a mixed wave tensor.†

The transformation formula for a covariant wave tensor has been found in (1·53), namely

$$S'=qS\bar{q}. \tag{7·21}$$

Let $q=e^{\frac{1}{2}E_\mu\theta}$. First let E_μ be an antisymmetrical matrix, so that

$$(E_\mu)_{\beta\alpha}=-(E_\mu)_{\alpha\beta}.$$

Then
$$\begin{aligned}(e^{\frac{1}{2}E_\mu\theta})_{\beta\alpha}&=\cos\tfrac{1}{2}\theta.(1)_{\beta\alpha}+\sin\tfrac{1}{2}\theta.(E_\mu)_{\beta\alpha}\\&=\cos\tfrac{1}{2}\theta.(1)_{\alpha\beta}-\sin\tfrac{1}{2}\theta.(E_\mu)_{\alpha\beta}\\&=(e^{-\frac{1}{2}E_\mu\theta})_{\alpha\beta}.\end{aligned}$$

That is to say,
$$\bar{q}=q^{-1}.$$

If E_μ is a symmetrical matrix, we have $\bar{q}=q$. Hence (7·21) becomes

$$\left.\begin{aligned}S'&=qSq^{-1}\text{ for antisymmetrical transformation matrices}\\&=qSq\text{ for symmetrical transformation matrices.}\end{aligned}\right\} \tag{7·22}$$

The transformation qSq^{-1} agrees with that of a mixed wave tensor (1·463). The transformation qSq has been called an antiperpendicular rotation (4·16). We have therefore the result:

A strain vector behaves as a space vector when the transformation matrix is antisymmetrical, but undergoes antiperpendicular rotations instead of the corresponding ordinary rotations when the transformation matrix is symmetrical.

As in the case of space vectors, we shall use the name "strain vector" indifferently for the array of components s_μ or for their symbolic combination $S=\Sigma s_\mu E_\mu$.

A strain vector may be represented graphically by plotting its components s_μ in a 16-dimensional space. But it must be remembered that the line which represents it will be thought of as a space vector. When a tensor

† This is the fundamental definition of a strain vector. It will be shown in § 7·6 that three-dimensional vector densities are strain vectors, and they have therefore been called strain vectors in anticipation (§ 4·6).

transformation is applied, all geometrical reference lines in the space rotate as space vectors; the strain vector follows its own transformation law, and is displaced relative to this background of space vectors. We shall now examine this relative displacement.

Let S be a strain vector and T a space vector, and at first let T coincide with S. Perform the infinitesimal transformation $q=e^{\frac{1}{2}d\Theta}$, where $d\Theta$ is a general infinitesimal matrix. We write $d\Theta=d\Theta_s+d\Theta_a$, where $d\Theta_s$, $d\Theta_a$ are its symmetrical and antisymmetrical parts. Since $d\Theta$ is infinitesimal, q is equivalent to the transformations $q_a=e^{\frac{1}{2}d\Theta_a}$ and $q_s=e^{\frac{1}{2}d\Theta_s}$ applied successively. Thus we have

$$S'=q_sq_aSq_a^{-1}q_s, \quad T'=q_sq_aTq_a^{-1}q_s^{-1}. \tag{7·23}$$

And, since $S=T$,

$$S'=T'q_s^2=T'e^{d\Theta_s}. \tag{7·24}$$

Hence, $d\Theta_s$ being infinitesimal,

$$S'-T'=T'd\Theta_s=Sd\Theta_s. \tag{7·25}$$

We take S to be non-singular. Then $Sd\Theta_s\neq0$ for any non-zero value of $d\Theta_s$. Since the general symmetrical matrix contains ten independent constants, the relative displacements $Sd\Theta_s$ will occupy a ten-dimensional space. We call this the *phase space* of the strain vector.

To exhibit this graphically, let S and T be represented by the lines OQ, OP in the 16-space. A tensor transformation of the space vector is represented by keeping OP fixed and rotating the axes of reference, thereby altering the components t_μ referred to the axes; thus P can be regarded as a fixed origin. At first Q coincides with P; but the transformation produces the relative displacement $PQ=S'-T'$. The interesting point is that, although there are 16 independent tensor transformations, Q is limited to a ten-dimensional locus. Provided that S is non-singular, every direction in this locus is a possible direction of PQ.

This construction gives only an infinitesimal region of phase space, and we must extend it by a process of continuation. The problems which arise in constructing the complete phase space will be considered later.

We have seen that antiperpendicular rotations represent intrinsic deformations of the physical system considered. The points of phase space therefore represent in systematic order different intrinsic states or configurations† of a system described by a strain vector. This corresponds to the usual definition of a phase space in statistical mechanics. In the applications for which it is intended, phase space is occupied by a probability distribution.

† I shall generally use the term *configuration*; it is to be understood in its broadest sense. *State* would express the meaning better; but I was anxious to avoid using a term which has been given a technical significance by Dirac. The term *phase* is often used in statistical theory; but I have reserved it for the angular variables occurring in the exponentials.

7·3. Real Phase Space.

For treating strain vectors and phase space we adopt a frame E_μ consisting of *four-point matrices*. The E_μ then consist of ten symmetrical and six antisymmetrical matrices; the symmetrical matrices are imaginary and the antisymmetrical matrices real. In § 6·1 space-like directions have been associated with imaginary matrices and time-like directions with real matrices; it is convenient to use these descriptions generally—as a nomenclature rather than as an anticipation of the physical manifestations of the vector components associated with them. Having regard to a future extension to double wave tensors, the definition is best stated in the form:

A *space-like* matrix is homothetic with its eigenvalue; a *time-like* matrix is antithetic to its eigenvalue.

The eigenvalues of the E_μ are imaginary. Thus, for simple wave tensors, we have the equivalence:

$$\text{Space-like} = \text{symmetrical} = \text{imaginary matrices},$$
$$\text{Time-like} = \text{antisymmetrical} = \text{real matrices}.$$

The ten dimensions of phase space are space-like.

We can no longer represent the E_μ by general fourfold matrices, since these are usually neither symmetrical nor antisymmetrical. At first sight the limitation to a particular frame of matrices seems a serious loss of "relativity". But the application of this chapter is to the internal wave tensors of a system, which, as we have seen, are not subject to Lorentz transformations. It is not merely permissible to use a fixed frame; it is essential that the frame to which these internal tensors are referred should have some quality which resists Lorentz transformations.‡

In treating external space we considered it to be a drawback that in matrix representation the E_μ have properties additional to those which they are defined to have as constituents of a complete orthogonal set. But the additional properties come in useful in internal space, because we have to indicate a distinctive direction, viz. that of the external momentum vector, and the planes of simultaneity orthogonal to it. We do this by giving to the matrix E_4, belonging to the distinctive direction, the property of antisymmetry; matrices belonging to directions in the plane of simultaneity are symmetrical; matrices belonging to intermediate directions are neither symmetrical nor antisymmetrical.

We have seen that one of the difficulties of applying to physics an algebra comprising complex numbers is that half the mathematical possibilities have to be set aside as unreal, i.e. not corresponding to actual phenomena. We approach the problem of determining the reality conditions of phase

† It is the Lorentz transformations of the frame which introduce asymmetry. For spatial rotations the matrices preserve their symmetry or antisymmetry.

space somewhat differently from the similar problem in Chapter VI. The feature of phase space is that it is the seat of a probability distribution. If the probability distribution does not extend to the unreal configurations—if they have zero probability—there is no need to stigmatise them further. Accordingly the discrimination of the world of real phenomena from the world of unreal phenomena is made when we insert the probability distribution. We shall show that this fixes the reality conditions of phase space unambiguously.

The method of statistical mechanics contemplates an initial or *a priori* probability distributed over phase space so that the probability in any volume is proportional to the volume. Observational information modifies this initial probability, so that the actual probability concerned in a particular problem is the product of the initial probability and a modifying factor. It is essential to the method that the phase space over which the probability is distributed should have finite volume; for if it were infinite, the initial probability associated with any finite region would be zero, and the method would break down. This requires that phase space should be a closed space.

It is true that statistical mechanics is often applied to a space closed by a supernatural barrier instead of by its own re-entrancy. But the barrier is merely a compression of the curvature required to close the space into a singularity. If we reverse the motion of the particles by natural fields of force instead of supernaturally, space must be curved to represent these fields. No ordinary field is such that an electron has zero probability of leaking through; to confine it rigorously the curvature must be sufficient to close the space.

Thus, in order that our phase space may be the seat of a probability distribution, it must be closed. This means that the matrix $d\Theta_s = \Sigma_s E_\mu d\theta_\mu$, which gives the displacements in phase space, must correspond to circular, not hyperbolic rotations. This requires that the $d\theta_\mu$ shall be real (cf. (4·15) and (6·13)). Then, since the symmetrical E_μ are imaginary, $d\Theta_s$ is imaginary. We have the result:

In order that the displacement in phase space given by the transformation $q = e^{\frac{1}{2}d\Theta_s}$ may represent a real change of configuration, $d\Theta_s$ must be an imaginary matrix. (7·31)

This reality condition may also be expressed by saying that the space is a phase space *in the other sense of the term phase*. Instead of a single algebraic phase θ indicated by a factor $e^{i\theta}$, we have ten phases for which i is replaced by different matrix roots of -1. Since the θ_μ are real, they are real periodic phase angles in the ordinary sense.

We can see this most easily by taking the initial value of the strain vector to be unity. This involves no loss of generality. By (7·24) $S' = T' e^{d\Theta_s}$. Since S' and T' are non-singular, T' has a reciprocal T'^{-1}; then

$$(T^{-1}S)' = T'^{-1}T' e^{d\Theta_s} = e^{d\Theta_s}.$$

The transformation law of $T^{-1}S$ is

$$(T^{-1}S)' = qT^{-1}q^{-1}.qS\bar{q} = qT^{-1}S\bar{q},$$

so that $T^{-1}S$ is a strain vector, and its initial value is 1. Thus the phase space generated by the transformations of an arbitrary non-singular strain vector S may equally be regarded as generated by the transformations of a unit strain vector, viz.

$$(1)' = 1.e^{d\Theta_s}. \tag{7·32}$$

If $d\Theta_s$ does not involve non-commuting components, (7·32) can be integrated so as to apply to finite displacements. Thus there will be a line of configurations for which the strain vector is

$$(1)' = e^{E_\mu\theta_\mu} = \cos\theta_\mu + E_\mu \sin\theta_\mu.$$

If θ_μ is real, the strain vector repeats itself at intervals $\theta_\mu = 2\pi$, so that we return to the original configuration. In other words phase space is re-entrant in the θ_μ direction. If θ_μ is imaginary, the strain vector is non-periodic and the phase space is open in that direction. Hence, in order that phase space may be closed, real configurations must correspond to real phase angles θ_μ.

Following an arbitrary track in phase space the transformations are in general non-commutative. Thus the exponentials combine by non-commutative multiplication, and the increments of the phases combine by non-commutative addition. The ten-fold phase is therefore non-integrable; this means that it must be represented in a curved space.

By § 3·6 the transformations $q = e^{\frac{1}{2}E_\mu d\theta_\mu}$, in which $d\theta_\mu$ is real, are all unitary. Thus a strain vector which is non-singular initially remains non-singular throughout phase space.

By (7·31) and § 6·5 (3) the matrix $d\Theta_s$, which represents displacement in phase space, satisfies the reality conditions for a strain vector, but not those for a space vector or vector density. We must therefore take $d\Theta_s$ to be a strain vector.

7·4. Coordinates in Phase Space.

Let $q = e^{\frac{1}{2}d\theta_s}$ be the transformation which displaces a point Q to a neighbouring point Q' in phase space, and let

$$d\Theta_s = \Sigma_s E_\mu \theta_\mu. \tag{7·41}$$

Then the ten θ_μ provide locally a coordinate system for describing points Q' near an origin Q. We call these *local orthogonal coordinates* or *natural coordinates*. A corresponding system of linear coordinates $x_\mu = R\theta_\mu$ is introduced by attributing to the phase space a definite scale constant R. For the present R must be considered arbitrary, since it could only be defined by introducing relations to an extraneous system.

The volume of a ten-dimensional element of phase space is defined to be

$$d\omega = d\theta_1 d\theta_2 \ldots d\theta_{10}. \tag{7·421}$$

It is a scalar quantity. We can, if we prefer, write it as $E_1 d\theta_1 . E_2 d\theta_2 \dots E_{10} d\theta_{10}$, but the product of the ten matrices is found to be ± 1. We thus have a definition of equal volumes at different points of phase space, each volume being measured in terms of natural angular coordinates at the point where it lies. By this definition *equivalent* volumes are *equal* volumes. As in § 2·9 equivalent volumes are formed by making the same construction in equivalent frames.

For any other system of coordinates x_μ, we have in the notation of general relativity

$$d\omega = \sqrt{-g} . d\tau, \tag{7·422}$$

where

$$d\tau = dx_1 dx_2 \dots dx_{10}.$$

In particular for natural linear coordinates

$$\sqrt{-g} = R^{-10}. \tag{7·43}$$

The local orthogonal system is only applicable when the squares of θ_μ are neglected. The most important equations of physics are differential equations of the second order, and in order to investigate them it is necessary to introduce a coordinate system valid at least as far as the squares of the coordinates. This problem will be treated in the next section. We have not much occasion to employ the properties of phase space as a whole, and our methods are chiefly adapted for treating an infinitesimal region. But it is important for our theory to prove that the whole volume of phase space is finite. Although this seems rather obvious, I have had some difficulty in proving it formally. We have adopted circular rotations in order to secure finiteness; but until we examine how they are to be extended beyond an infinitesimal region we cannot be sure that they will achieve this end. They secure re-entrancy along geodesic tracks, but we have still to prove that phase space has no tortuous exit. Although phase space has the same kind of uniformity as a hypersphere, it is different from a hypersphere; it contains pairs of antiperpendicular directions. Thus ordinary spherical coordinates do not apply, and I do not know of any suitable adaptation of them.

There is a fundamental difficulty in specifying finite deformations or strains of a system, which arises in the following way. Let A_1, A_2 denote two different orientations of the same (unstrained) system, and let A' denote the system in a strained condition; how are we to decide whether the strain should be measured by $A' - A_1$ or $A' - A_2$? There is no absolute one-to-one correspondence of the orientations of strained and unstrained systems; but it is necessary to lay down some conventional rule which will prevent our representing the same deformation twice over as $A' - A_1$ and $A' - A_2$.

We have secured a unique representation of infinitesimal strains. The initial (unstrained) state was represented by S or T; the strained state S'

was then compared with T', which represents the unstrained state in a different orientation connected with S' by tensor rules. This cannot be extended to finite regions, because the transformations $T \to T'$ are not integrable. To meet this difficulty, consider the transformation

$$q = \Pi q_\mu = \Pi e^{\frac{1}{2}E_\mu \phi_\mu}, \tag{7·44}$$

the product consisting of ten factors with space-like E_μ arranged in a fixed order. The elementary transformations q_μ are applied successively (in the reverse order to that in which they are written). The ϕ_μ are not infinitesimal. Applying this transformation to a strain vector, so that $S \to S'$, we obtain a strained configuration which will correspond to some point Q' in phase space. We can adopt ϕ_μ as the coordinates of Q'.

Let one of the ϕ_μ, say ϕ_σ, receive an increment $d\phi_\sigma$. Let the new point be Q'', and the new value of q be q''. For two matrices q, q'' which differ infinitesimally we can find a matrix $d\Theta$ such that

$$q'' = e^{\frac{1}{2}d\Theta} q.$$

It is easily seen that

$$d\Theta = X E_\sigma d\phi_\sigma X^{-1}, \tag{7·45}$$

where X is the part of the product Π which precedes q_σ. X cannot be singular (§ 3·6). In general $d\Theta$ will include time-like matrices. Since $e^{\frac{1}{2}d\Theta}$ is the transformation which changes Q' to Q'', the components of $d\Theta_s$ are the natural coordinates $d\theta_\mu$ of Q'' referred to the origin Q'.

Similarly, we can express the other displacements $d\phi_\mu$ in natural coordinates $d\theta_\mu$ at Q'. The jacobian $\partial(\phi_\mu)/\partial(\theta_\mu)$ gives the ratio of the volume element $d\tau = d\phi_1 d\phi_2 \dots d\phi_{10}$ to the natural volume element $d\omega = d\theta_1 d\theta_2 \dots d\theta_{10}$ and hence determines $\sqrt{-g}$ in the usual formula $d\omega = \sqrt{-g}\, d\tau$.

The half-period of each of the ϕ_μ is 2π. After each half-period the values of q repeat themselves with opposite sign, and the corresponding strain vectors repeat themselves; so that the half-period represents a circuit of phase space. Hence for the whole of phase space, or of that part of phase space covered by the coordinate system ϕ_μ,

$$\int d\tau = (2\pi)^{10}.$$

Thus unless $\sqrt{-g}$ becomes infinite anywhere, or unless there are configurations not obtainable by the transformation (7·44), the volume $\int d\omega$ is finite.

Since X cannot be singular, (7·45) shows that $\partial\theta_\mu/\partial\phi_\sigma$ is not infinite, and hence $\sqrt{-g}$ is never infinite. There are, however, loci where $\sqrt{-g} = 0$. For example, if the first two factors of Π are $e^{\frac{1}{2}E_\nu\phi_\nu} e^{\frac{1}{2}E_\sigma\phi_\sigma}$, where E_ν, E_σ anticommute, we have by (7·45) for a displacement $d\phi_\sigma$

$$d\Theta = e^{\frac{1}{2}E_\nu\phi_\nu} E_\sigma d\phi_\sigma e^{-\frac{1}{2}E_\nu\phi_\nu} = e^{E_\nu\phi_\nu} E_\sigma d\phi_\sigma.$$

On the locus $\phi_\nu = \frac{1}{2}\pi$, this becomes $E_\nu E_\sigma d\phi_\sigma$. The product of two anticommuting space-like matrices is a time-like matrix; thus $d\Theta$ is wholly

time-like, and $d\Theta_s$ vanishes.† But this type of singularity, in which a small natural volume is infinitely magnified in coordinate volume, is harmless for our purpose.

I have little doubt that a geometer could furnish a more elegant proof. Probably he could evaluate the volume. But I think the above investigation satisfies us that the volume of phase space is finite.

7·5. Stereographic Coordinates.

One of the most important practical steps in the theory is to provide the analytical machinery for investigating differential equations of the second order. For this purpose we introduce a system of coordinates valid (under certain restrictions) for finite regions of phase space.

On a hypersphere of radius R, stereographic coordinates are such that the line element is

$$ds^2 = -(1+r^2/4R^2)^{-2}(dx^2+dy^2+dz^2+\ldots), \tag{7·51}$$

where $r^2 = x^2+y^2+z^2+\ldots$. The coordinates are thus locally orthogonal and isotropic, but not uniform; the actual length ids is λ_r times the Euclidean length $(dx^2+dy^2+dz^2+\ldots)^{\frac{1}{2}}$, where

$$\lambda_r = (1+r^2/4R^2)^{-1}. \tag{7·52}$$

The hypersphere is in this way projected into a Euclidean space with a variable gauge factor λ_r.

Analogous coordinates x_μ in phase space are defined as follows. Let $X = \Sigma_s E_\mu x_\mu$, the summation being restricted to the space-like matrices. Then stereographic coordinates are such that the displacement from the origin to the point x_μ corresponds to the transformation

$$q = \left\{\frac{1+X/2R}{1-X/2R}\right\}^{\frac{1}{2}}. \tag{7·53}$$

The right-hand side is to be interpreted by expanding in infinite series. This formula is limited to a domain containing only perpendicular coordinates together with its infinitesimal neighbourhood in all directions. That is to say, X is limited to a pentadic expression or to a single algebraic variable, but dX is unrestricted. The inclusion of the infinitesimal neighbourhood is essential, because these coordinates are used principally when we are treating the complete ten-dimensional volume element of phase space.

Accordingly X^2 is algebraic,‡ and we set

$$X^2 = -r^2. \tag{7·54}$$

† This is illustrated graphically by taking E_σ, E_ν to correspond to rotations of a sphere in the planes xy, yz. Taking an origin on the x axis, a displacement from the origin to any point on the sphere can be represented by two such rotations in the order given. Different values of ϕ_σ, ϕ_ν will give different points, unless $\phi_\nu = \frac{1}{2}\pi$ or $\phi_\sigma = 0$. This example shows that the singularity of the representation at these two points does not signify a failure of the coordinate system ϕ_μ to cover the whole of the space.

‡ It is tempting to describe the domain of X as being limited by the condition that X^2 is algebraic; but we define it more stringently in order to exclude compact E-numbers (5·66).

If displacement from x_μ to $x_\mu+dx_\mu$ corresponds to the transformation $e^{\frac{1}{2}d\Theta}$, we have by (7·53)

$$e^{\frac{1}{2}d\Theta}\left\{\frac{1+X/2R}{1-X/2R}\right\}^{\frac{1}{2}}=\left\{\frac{1+(X+dX)/2R}{1-(X+dX)/2R}\right\}^{\frac{1}{2}}. \tag{7·55}$$

When X is limited as above, the most general matrix dX can be divided into two parts, one of which commutes and the other anticommutes with X; for the limitation secures that a complete set can be formed with X/r as one of its members; and if dX is resolved in that frame, its components either commute or anticommute with X/r. The transformations corresponding to the two parts of dX can be treated successively, since they are infinitesimal.

First, let dX commute with X. Then, since there are no non-commutative symbols, (7·55) can be solved like an algebraic equation, giving

$$d\Theta=d\left\{\log\frac{1+X/2R}{1-X/2R}\right\}=(1-X^2/4R^2)^{-1}dX/R. \tag{7·56}$$

Next, let dX anticommute with X. Then the differential of any even power of X is zero. Since

$$\left\{\frac{1+X/2R}{1-X/2R}\right\}^{\frac{1}{2}}=\frac{1+X/2R}{(1-X^2/4R^2)^{\frac{1}{2}}},$$

$$d\left\{\frac{1+X/2R}{1-X/2R}\right\}^{\frac{1}{2}}=\frac{dX/2R}{(1-X^2/4R^2)^{\frac{1}{2}}}=\frac{dX}{2R}\left\{1+\frac{X}{2R}\right\}^{-1}\left\{\frac{1+X/2R}{1-X/2R}\right\}^{\frac{1}{2}}.$$

Hence by (7·55)

$$e^{\frac{1}{2}d\Theta}=1+\frac{dX}{2R}\left\{1+\frac{X}{2R}\right\}^{-1}=1+\frac{dX}{2R}\left\{1+\frac{r^2}{4R^2}\right\}^{-1}\left\{1-\frac{X}{2R}\right\},$$

$$d\Theta=\left\{1+\frac{r^2}{4R^2}\right\}^{-1}\left\{\frac{dX}{R}-\frac{dX.X}{2R^2}\right\}. \tag{7·57}$$

By definition X and dX contain only space-like matrices, and since they anticommute their product is time-like. Thus (7·56) and (7·57) give the same value of $d\Theta_s$

$$d\Theta_s=(1+r^2/4R^2)^{-1}dX/R=\lambda_r dX/R \tag{7·58}$$

by (7·52).

Thus the differentials dx_μ are the natural linear coordinates at the point considered, but the scale constant R/λ_r is variable—precisely as in (7·51).

The volume of a ten-dimensional element being

$$d\omega=d\theta_1 d\theta_2\ldots d\theta_{10}=\sqrt{-g}\,dx_1dx_2\ldots dx_{10},$$

we have by (7·58)

$$\sqrt{-g}=(\lambda_r/R)^{10}=R^{-10}(1+r^2/4R^2)^{-10}. \tag{7·59}$$

The following theorem is required later:

If a transformation X (not infinitesimal, but not containing antiperpendicular components) is applied to the strain vectors, the stereographic coordinates of all points in the infinitesimal neighbourhood of the origin are changed by the same amount, to the first order.

Let the stereographic coordinates 0, dx_μ of the origin and a neighbouring point be changed by the transformation to $x_\mu, x_\mu + dx_\mu'$; and let $dX = \Sigma E_\mu dx_\mu$, etc. We have to show that $dX' = dX$. The transformation is that given by (7·53), and we have

$$\left\{\frac{1+X/2R}{1-X/2R}\right\}^{\frac{1}{2}} e^{dX/2R} = \left\{\frac{1+(X+dX'')/2R}{1-(X+dX'')/2R}\right\}^{\frac{1}{2}},$$

where dX'' differs (if at all) from dX' by including time-like components.† By the conditions imposed on X, we can divide dX into two parts which respectively commute and anticommute with X, and treat them separately. For the commuting part the left-hand side can be treated algebraically, and gives immediately the required result $dX'' = dX$. For the anticommuting part we proceed as in obtaining (7·57) and find that dX'' differs from dX by time-like components only, so that $dX' = dX$.

This result may also be stated in the form: In stereographic coordinates, a finite displacement of the above restricted type commutes with all infinitesimal displacements.

The results of this section are used extensively in Chapter XII.

7·6. Associated Strain Vectors and Space Vectors.

We employ four-point matrices, E_4, E_5 being as usual the real members of a pentad. Let e_μ be a covariant wave tensor which has the value E_μ in the coordinate system initially chosen. Then, after a transformation q, we have $e_\mu = qE_\mu q^{-1}$ or $qE_\mu q$ according as q contains a time-like or a space-like matrix (7·22).

The wave tensor e_{45} has the remarkable property that it is invariant for the ten rotations in five dimensions (kinematic rotations). This is easily verified, remembering that by the above formulae it is unaltered by transformations with real matrices with which it commutes or imaginary matrices with which it anticommutes. Thus for all orientations of the axes in five dimensions

$$e_{45} = E_{45}. \tag{7·61}$$

No other strain vector has this property. It may be compared with the metrical tensor $g_\mu{}^\nu = \delta_\mu{}^\nu$ and the contravariant tensor density $\epsilon_{\mu\nu\sigma\tau}$ in ordinary tensor calculus, which are likewise exceptional in having invariant values.

In ordinary tensor calculus we define *associated* covariant and contravariant vectors A_α, A^α by the relation $A_\beta = g_{\alpha\beta}A^\alpha$. We shall now define associated (initial) covariant and contravariant wave vectors ϕ^*, χ^*. A linear relation between them must be of the form

$$\phi_\beta = \chi^\alpha a_{\alpha\beta},$$

† The left-side gives the transformation $0 \to dx_\mu$ followed by the transformation $0 \to x_\mu$. By definition this takes us to the point $x_\mu + dx_\mu'$. We cannot immediately identify it with the direct transformation $0 \to x_\mu + dx_\mu'$, since an infinitesimal time-like matrix may be included.

where $a_{\alpha\beta}$ is a covariant wave tensor. Consider the special case $a_{\alpha\beta}=i\,(e_{45})_{\alpha\beta}$. Then $\phi_\beta=i\chi^\alpha(E_{45})_{\alpha\beta}$, by (7·61); or, dropping suffixes,

$$\phi^*=i\chi^*E_{45}. \tag{7·62}$$

Wave vectors connected by the relation (7·62) will be called *associated wave vectors*.

Thus in wave-tensor calculus, $(iE_{45})_{\alpha\beta}$ plays the part of $g_{\alpha\beta}$ in the operation of lowering a suffix.† Since $(iE_{45})^2=1$, we have also $\chi^*=i\phi^*E_{45}$, which defines the operation of raising a suffix. Since four-point matrices are used, raising or lowering a suffix is a rather simple process; for example, in the standard pentad (3·27),

$$E_{45}=\begin{matrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{matrix}$$

Hence
$$\phi_1,\ \phi_2,\ \phi_3,\ \phi_4=i\,(\chi^3,\ \chi^4,\ -\chi^1,\ -\chi^2). \tag{7·63}$$

Now multiply (7·62) by a covariant wave vector ψ by outer multiplication. We have

$$\psi\phi^*=i\psi\chi^*E_{45}.$$

Or, denoting the covariant wave tensor $\psi\phi^*$ by S, and the mixed wave tensor $\psi\chi^*$ by T,

$$S=iTE_{45}. \tag{7·64}$$

We have therefore the important result:

If a space vector T is multiplied by iE_{45} we obtain an associated strain vector S. Reciprocally, if a strain vector S is multiplied by iE_{45} we obtain an associated space vector T.

The components of S are obtained by shuffling the components of T, and in some cases inserting factors $\pm i$; the process is thus somewhat analogous to (7·63). We have already given the precise relation between the s_μ and t_μ in (4·66).

The relation (7·64) is invariant for kinematical rotations, but electrical rotations (§ 6·4) are excluded. It would be undesirable to exclude electrical rotations of a strain vector; because a strain vector is used primarily in connection with phase space, and some of the directions of displacement in phase space correspond to electrical matrices. But we have suggested that the term "space vector" implies that only kinematical rotations are contemplated (§ 4·3). It would appear therefore that the strain vector is the more fundamental conception, and that space vectors are a derivative conception introduced into physics by the formula (7·64). That is to say, we take S to be a covariant wave tensor for *all* transformations; then the

† If we could depend on the suffixes appearing explicitly in the formulae, we should naturally use the notation χ^a, χ_a for the two associated wave vectors; but since suffixes are generally omitted, we have to distinguish them by different letters χ, ϕ.

matrix T determined from it by (7·64) will behave as a mixed wave tensor for kinematical rotations but not for electrical rotations; so that it will be rigorously a space vector, but only imperfectly a mixed wave tensor. If all space vectors in physics originate out of strain vectors in this way, we see why there is never any occasion to employ in practice the extra relativity transformations possessed by space vectors derived from mixed wave tensors. We have previously attributed the absence of these transformations to the neutrality of space-time. The two explanations are connected; because a system composed of particles of one sign could have no equilibrium configuration, and give no foothold for statistical mechanics with its attendant conception of phase space.

Since (7·64) is the same equation as (4·65), S is the three-dimensional density of T; and we reach our earlier definition of a strain vector as the three-dimensional density of a space vector. The relation is reciprocal, and T is also the three-dimensional density of S. But the discussion in § 4·6 was limited to an infinitesimal region round the origin, where the volume element of the 3-space was $dW_{123}=iE_{45}dw$. If we move away from the origin, the matrix of dW_{123} will no longer be the original E_{45}. In addition to the change of direction of the normal to space-time, we must allow for an arbitrary rotation of the time direction, since there is no absolute way of defining the reckoning of simultaneity over an extended area. The general formula for the volume element is

$$dW_{123}=i\,(e_{45})_\alpha{}^\beta\,dw, \tag{7·65}$$

where $(e_{45})_\alpha{}^\beta$ is the space vector which has the value E_{45} in the original coordinate system at the origin. Hence the three-dimensional density of T is $iT\,(e_{45})_\alpha{}^\beta$, whereas the strain vector is $iTE_{45}=iT\,(e_{45})_{\alpha\beta}$.

We see therefore that the elementary definition of a strain vector as the three-dimensional density of a space vector, and *vice versa*, does not hold for an extended curved region. It can, however, be preserved if we represent the element of volume as a strain vector with matrix $(e_{45})_{\alpha\beta}$ instead of in the more familiar way as a space vector with matrix $(e_{45})_\alpha{}^\beta$. To regularise this we distinguish an internal and an external three-dimensional space. They consist of the same points; a particle which is in one is in the other; but the metrical conceptions are different. The internal space is part of phase space, corresponding to three of its ten dimensions. We have seen (7·421) that the whole volume element is scalar; the separation of the three and the seven dimensions is so drawn that each combination is a strain vector with matrix E_{45}, constant throughout phase space, and therefore in a sense characteristic of the whole phase space.

The phase space, including the internal space, is described wholly by strain vectors. It will be remembered that the matrix $d\Theta_s$, determining displacement in it, was found to be a strain vector (§ 7·3). But when the

three dimensions are separated from the others, the densities or fluxes of strain vectors with respect to the three-dimensional volumes constitute space vectors. These require for their representation an external space.

7·7. Normalised Strain Vectors.

We shall now transform the wave equation so that it applies to strain vectors. The equations giving the factors ψ, χ^* of a space vector are

$$H\psi = 0, \quad \chi^* H = 0, \tag{7·71}$$

where $H = E_{15}p_1 + E_{25}p_2 + E_{35}p_3 + E_{45}p_4 - m$, by (6·51). Let

$$H_s = -E_{45}H = E_{14}p_1 + E_{24}p_2 + E_{34}p_3 + p_4 + E_{45}m. \tag{7·72}$$

Then (7·71) becomes $E_{45}H_s\psi = 0, \quad \chi^* E_{45}H_s = 0,$

so that

$$H_s\psi = 0, \quad \phi^* H_s = 0, \tag{7·73}$$

where $\phi^* = i\chi^* E_{45}$ by (7·62). These are the wave equations for the factorisation of a strain vector $\psi\phi^*$. The hamiltonian H_s is part of $\psi\phi^*$. If $p_4 = ip_0$, we have

$$H_s = E_{14}p_1 + E_{24}p_2 + E_{34}p_3 + E_{16}p_0 + E_{45}m. \tag{7·74}$$

This satisfies the reality conditions for a strain vector (§ 6·5 (3)), the coefficients in (7·74) all being real. The energy p_0 is associated with the algebraic matrix E_{16}; and the coordinate conjugate to it, namely the time t, must also be associated with E_{16}. *Accordingly in phase space the algebraic phase represents the time.*

Since the phase space corresponds to the internal configurations of a system, the time θ_{16} is measured in the direction defined by the external momentum vector (§ 7·1).

In phase space the algebraic phase may be separated from the other phases, leaving a nine-dimensional space. This is permissible, because the algebraic phase commutes with all the others. If $d\omega_c$ is the nine-dimensional volume element and $d\theta_{16}$ the algebraic phase, so that

$$d\omega = d\omega_c d\theta_{16}, \tag{7·751}$$

we can treat $d\theta_{16}$ separately in integrating; so that, if Ω is the volume of the ten-dimensional phase space, and Ω_c the volume of the nine-dimensional space,

$$\Omega = \Omega_c . 2\pi. \tag{7·752}$$

In defining phase space we associated a strain vector S with each point. We shall now more definitely associate a strain vector $\Sigma = S d\omega_c/\Omega_c$ with a range of configurations $d\omega_c$. This strain vector will serve three purposes:

(1) Its algebraic phase indicates the time.

(2) Its symbolic phases describe a particular configuration.

(3) Its amplitude indicates the probability of the system having this configuration to within a range $d\omega_c$.

This interpretation applies differentially to small changes of Σ, but it is not so easy to interpret Σ itself. For the latter purpose we must introduce the determinant of Σ. Since $\det \Sigma$ is invariant for all non-algebraic displacements in phase space (§ 3·6) it is independent of (2). For a purely algebraic strain vector, $\det \Sigma = \Sigma^4$. If then we take as "origin" the configuration for which the strain vector is algebraic (regarding this as the standard unstrained condition), we have

(1) The time is the argument of the complex number $(\det \Sigma)^{\frac{1}{4}}$.

(2) The configuration is specified by the matrix $\Sigma/(\det \Sigma)^{\frac{1}{4}}$.

(3) The probability is the modulus $|(\det \Sigma)^{\frac{1}{4}}|$.

It is a most important feature of the symbolic theory that *the same symbol specifies both the configuration and the probability that the system has that configuration.*

The probability distribution which we contemplate initially is uniform throughout phase space, so that an element $d\omega_c$ contains probability $d\omega_c/\Omega_c$. Hence for the initial probability $|(\det S)^{\frac{1}{4}}| = 1$. In particular, at the origin (standard unstrained configuration) $S = 1$. This initial or *a priori* distribution is the framework—the "blank sheet"—into which we insert whatever we may learn about the system by special observation. If the observational evidence shows that at a time t certain configurations were more probable than others, we inscribe on the blank sheet a function $f(\theta_1, \theta_2, \ldots t)$ indicating that the actual probability of the configuration $(\theta_1, \theta_2, \ldots)$ was f times the initial probability. We call f the *modifying factor*. The modified, i.e. the actual, distribution is therefore represented by a strain vector S which does not in general satisfy $|(\det S)^{\frac{1}{4}}| = 1$. The factor f is necessarily an algebraic function. In wave analysis it is expressed as the product of two wave functions, and is therefore formally a wave tensor of the second rank in which all terms are zero except the quarterspur. This suggests a generalisation of the modifying factor f. It often happens that in introducing the modified probability we make at the same time a transformation of coordinates; that is to say, we compare the modified probability of a configuration with the initial probability, not of the same configuration, but of a configuration related to it by a transformation. The transformation and the modification of probability are comprised in a non-algebraic modifying factor f, which is the product of vector wave functions.

Whether we are treating the initial or the modified distribution, its strain vector is normalised so that the total probability in the volume Ω_c is 1. In this normalisation the time dimension θ_{16} is excluded, because the conception of distributing probability over extension in time is rather unusual. An atom exists continuously in time, so that the association of its probability with particular time intervals dt does not arise in a direct way.

But consider a clock with one hand moving at a uniform rate. If the clock is part of the system, we shall, in specifying the configuration of the system, specify the position angle θ_{16} of the hand of the clock; and the elementary range of configurations $d\omega$ will include the element $d\theta_{16}$ of position angle. Thus from the internal point of view θ_{16} is an ordinary angular coordinate in the specification of configuration, and is one of the dimensions of the domain over which the unit probability is distributed. It may be included in the general normalisation; or (owing to its commutation with all the other phases) it may be excluded and normalised separately, so that the probability $d\theta_{16}/2\pi$ of a range $d\theta_{16}$ is stated separately.

It was rather surprising to find *time* appearing at all in phase space. We apparently rid ourselves of it when we retained only the displacements associated with space-like matrices. But it has gained entry in space-like disguise as the position angle of the hand of a clock. Moreover, it is a periodic angular coordinate, not an open infinite coordinate like external time.

This may be made clearer by an illustration from celestial mechanics. The orbit of a planet is specified by six elements, one of which is the epoch of perihelion passage T. The element T corresponds to θ_{16}. If there are several planets, we must include the element T for each of them in enumerating the possible systems which might be formed. Thus T is an essential coordinate of the configuration space in which we represent the possible combinations, although it is ordinarily conceived as having a time-like character.

Since time is measured by a phase angle, instants which differ by multiples of a period are to be considered identical; and the whole extent of time is $2\pi R$ in natural linear coordinates. That is because in the structure assigned to the system there is no provision for a revolution counter. A more extended time reckoning can only be given a meaning when we treat more complicated systems. It is fairly obvious that infinite time will appear automatically when we introduce systems with incommensurable periods.

7·8. Physical Meaning of Strain Vectors.

By the aid of space vectors we have defined a domain which has the primitive property of conceptual space-time, namely that all points of it are equivalent. It is all one whether a given object is at the point A or at the point B. But this is in flagrant contradiction to our experience that somehow it is possible to find out that the object is at A, not B. We have to combine two different conceptions of position—absolute and relative. By *absolute* I mean "conceived as absolute", i.e. pictured in an abstract geometrical frame; by *relative* I mean "relative to observable landmarks". In so far as the points of space-time symbolise absolute positions they are equivalent to one another; in so far as they symbolise relative positions they can be dis-

criminated observationally. According to whether the point is being considered in its absolute rôle or its relative rôle, the position vector defining it is a space vector or a strain vector.

The same applies to other physical entities. The energy and momentum constitute a vector. Referred to an abstract geometrical frame, this is a space vector, but relative to another physical system it is a strain vector. The difference between space vectors and strain vectors might be defined by saying that when we treat of space vectors we are contemplating the "blank sheet"; when we treat of strain vectors we are beginning to write something on the clean page.

It is true that when we contemplate *two* space vectors we are implicitly writing something on the clean page, namely the internal relations of a system consisting of two parts. But the treatment of two vectors, and the extraction of the internal relations of that which they represent, is slightly more advanced than the problem we are now handling, and is not suitable for our first writing lesson. We here treat the internal relations as already extracted and presented to us in the form of a strain vector.

The conception of relative position or relative momentum arises when, instead of contemplating the particle as a solitary system in an abstract frame, we regard it as part of a more comprehensive physical system. Consider, for example, an electron in an atom. A displacement of the electron has two aspects. It is a translation of the electron from one point to another in external space-time; as such it is a change of the space vector defining the absolute position of the electron. But it is also a deformation or strain of the atom; as such it is a change of a strain vector. The atom is changed to a new configuration, so that there is a displacement in the phase space in which the configurations of the atom are represented.

The tensor calculus provides a machinery for locking the changes of one characteristic to those of another. The changes of strain of the system are locked to the changes of absolute position of its parts. We should commonly say that the change of position of the electron is the *cause* of the change of strain of the atom. To work out this connection in an actual system of several particles is too complex a problem for us at present. We proceed in the converse way. We consider the simplest form of tensor interlocking, and construct the ideal "system" to which it would apply. We cannot find any more elementary starting point than the locking of a covariant wave tensor to a mixed wave tensor; and the former has been called a strain vector in anticipation of this application.

Accordingly when we wish to pass from the absolute to the relative aspect of the position of a particle, we treat the particle as an element of an ideal physical system whose configurations (so far as they are determined by

this particle) are specified by a simple strain vector, and therefore occupy a phase space of the kind we have been investigating.

What exactly has been added to the particle to make it part of a physical system instead of a lone particle—to de-absolutise it? Mathematically it is the matrix E_{45}, which transforms its space vector into its strain vector. This matrix defines a particular three-dimensional section of the world—a plane of simultaneity. Physically the particle is made part of a system by associating with it a plane of simultaneity.

We have seen that when a number of particles are treated as a combined system, each particle retains its separate space coordinates, but there is only one time coordinate for the whole system. This replacement of the individual time coordinates by a common time coordinate is the essence of the process of *combining*; it defines the change in our point of view when we consider the system as a whole instead of its constituent parts. A hydrogen atom is composed of a proton and electron; but a proton today and an electron yesterday do not constitute a hydrogen atom. We have seen (§ 7·1) that the planes of simultaneity which correspond to the common time coordinate of the system are determined by the direction of its external momentum vector.

We can now see how the conversion of space vectors into strain vectors corresponds to our change of attitude when we consider the particle to be part of a system. Actually to introduce the other particles of the system would greatly complicate the problem; but, in anticipation of their presence, we introduce the planes of simultaneity which will be determined by the external momentum vector of the system when it appears. We construct these planes in the geometrical space-time which previously contained no indication of a particular direction of section; we inscribe the matrix E_{45}, henceforth to be permanently associated with the planes of simultaneity, on the sheet which was previously blank. Simultaneity is no longer arbitrary; we cannot modify the reckoning of it to suit a particular particle of the system. Suppose that the particle has a momentum in the x-direction; by (6·51) the space vector representing this is $E_{15}p_1$. Formerly the particle could be "reduced to rest" by a Lorentz transformation with matrix of the form $E_{14}iu_{14}$; but we are no longer allowed to rotate the plane of simultaneity. The strain of the system (as compared with a system in which the particle is at rest) is measured by the inhibited rotation $E_{14}iu_{14}$. Apart from a numerical factor this is the strain vector $E_{15}p_1 . iE_{45} = E_{14}ip_1$ associated with the space vector $E_{15}p_1$.

This example calls attention to another feature of the connection between internal and external space. The external velocity or momentum vectors correspond to internal displacement vectors, and *vice versa*. At present we can only recognise this in a preliminary way; but another example may be

of interest. If we apply this principle to the internal energy attached to the matrix E_{16}, the corresponding space vector attached to the matrix E_{45} should be interpretable as an external time displacement. As in the previous example it will be an inhibited or virtual time displacement. If we change the scale of the system, we change the light-times between its various parts, and therefore change the time at which a particle becomes causally effective at the centroid where the external wave vector is supposed to be located. But in our internal wave functions the particles are assigned simultaneous times, not the times at which they are causally effective. *Thus the distance of a particle from the centroid can be looked upon as an inhibited time displacement.* As an isolated particle we should have contemplated it at an antedated instant to allow for the lag of causal efficiency; as part of a system we can only contemplate it at the same instant as the rest of the system. If it receives a radial displacement δr, a further lag $\delta r/c$ is introduced which (if the particle were not considered to be part of the system) would be compensated by giving it a time displacement $\delta r/c$. Since the planes of simultaneity in the system are fixed, we cannot give an individual time displacement to the particle; and the inhibited time displacement appears as, and is the measure of, a strain of the system. We have seen that this strain will be associated with the matrix E_{16} and therefore be an internal energy. Thus a system will in general have an internal energy depending on its linear scale. This is one aspect of the origin of electrostatic energy.†

Our present point of view is that, instead of starting with an elementary particle defined to be such that its properties can be represented by a complete space vector, it is rather less abstract to start with an elementary system defined to be such that its condition or "configuration" can be represented by a simple strain vector. Then by introducing the associated space vector we detach the active principle of the system from the passive principle represented by planes of simultaneity, and so obtain a still more abstract entity, namely an elementary particle in free space. It may seem far-fetched to describe a particle coupled with a plane of simultaneity as a "physical system". But that is as much of a physical system as we can represent by a simple wave tensor. And inasmuch as current quantum theory has made shift to treat a wide range of problems of observational importance with simple wave tensors, it is an important stage in the advance of the theory towards actuality.

If a system A were completely isolated it would be unobservable. Therefore it enters into observation as part of a more extended system B. As such it will be represented in the internal space of B. The external momentum

† Conversely, if we derive the electrostatic energy as in Chapter xv, the foregoing investigation shows how to connect with it the idea of antedating the particles by the light-time.

vector of B determines planes of simultaneity for the internal constituents of B; and the external momentum vector of A, which was a space vector so long as A was considered in isolation, is replaced by the associated strain vector when A is considered as part of B. But B in its turn would be unobservable if it were not part of a larger system C; its external momentum vector must therefore be replaced by a strain vector, representing B as a constituent of the internal structure of C. And so on.

At each stage there is an external space vector of the system contemplated, which is converted into a strain vector when we consider a larger system—until the system has been extended to comprise the whole universe, when the external vector can be dropped, all phenomena being comprehended in the internal structure.

From this point of view space vectors make only a temporary appearance —when we halt for breath in the course of the analysis. As soon as we are ready to proceed further, we replace them by their associated strain vectors. There are, however, two reasons why space vectors remain an important conception in practical investigation. Firstly, it is impracticable to treat the exact equations of the internal states of highly complex systems; therefore when we have to deal with a number of systems with weak interaction, we leave them uncombined and treat their mutual influence (which brings them within reach of observation) by approximate perturbation methods. Secondly, in dealing with a large number of similar systems, we combine them *statistically*, not *individually*. The external space vectors are not transformed individually into internal strain vectors of the complex system; but are first replaced by a probability or average distribution. The internal state of the complex system is described by a very much simplified set of strain vectors embodying coefficients of the distribution function. Thus the internal state of a gas is described by pressure, energy density, virial, vorticity, etc., representing certain averaged characteristics of the external space vectors of the molecules. Normally the direct procedure of replacing space vectors by their associated strain vectors is not extended to systems greater than a molecule.†

7·9. Singular Phase Space.

The phase space which we have been considering is generated by the transformations of a non-singular strain vector. If we use instead a singular strain vector S_a, the resulting phase space has fewer dimensions, since $S_a d\Theta_s = 0$ when $d\Theta_s$ is a pseudo-reciprocal of S_a. Displacement in such a direction involves no change of the strain vector and therefore no change of

† The averaging could perhaps equally well be performed on the associated strain vectors. But since space vectors are more familiar we generally adhere to them as long as we can.

configuration; the direction accordingly is not a dimension of the phase space of S_a.

To show the relation between singular and non-singular phase space, we consider a non-singular strain vector which at some point of phase space has a purely algebraic value S_0. By (6·64) this is analysed into spectral components

$$S_a + S_b + S_c + S_d = S_0. \tag{7·91}$$

The components are pure and therefore singular. Consider the component S_a. Since the products $S_a S_b$, etc. are zero, the singular directions, which satisfy $S_a d\Theta_s = 0$, are

$$d\Theta_s = bS_b + cS_c + dS_d. \tag{7·92}$$

Since $d\Theta_s$ cannot contain time-like matrices, the ratio $b : c : d$ must be chosen so as to eliminate the time-like matrices E_{23}, E_{45}. There is therefore only one singular direction

$$d\Theta_s = c\,(S_c + S_d). \tag{7·93}$$

If, as in (6·64),

$$S_a = -\tfrac{1}{4}iS_0\,(E_1 + E_{23} + E_{45} + E_{16}), \tag{7·941}$$

the singular direction is

$$d\Theta_s = (-E_1 + E_{16})\,d\phi. \tag{7·942}$$

For S_c and S_d the singular direction is $d\Theta_s = (E_1 + E_{16})\,d\phi$.

Non-singular phase space of ten dimensions can thus be regarded as a superposition of two singular phase spaces of opposite spin, each of nine dimensions. The axis of spin, here represented by E_1, can be in any direction in three-dimensional space. When the singular phase space of S_a is delineated in ordinary phase space, the added dimension (direction of ϕ) gives a line of indistinguishable configurations which are really one configuration. The equation of the singular line (7·942) may be written

$$d\phi = -d\theta_1 = d\theta_{16}.$$

Or, since $dt = R\,d\theta_{16}$,

$$d\theta_1/dt = -1/R.$$

The singular line therefore represents an entity spinning with uniform velocity. We do not distinguish the different orientations corresponding to the sequence of points along the line, but count the state of spin as one configuration—a constant state of strain of the system. The strain produced by the rotation is to be compared with the gyrostatic torque of a fly-wheel, not with the torque of a wound-up spring. A change of the plane of spin would be a change of configuration.

When we separate non-singular phase space into two singular phase spaces, we require an additional variable to specify how the probability is divided between the two phase spaces. This extra variable compensates for the loss of a dimension in passing to singular phase space. It is to be remembered that $S_a + S_c$ does not represent two particles of opposite spin; two particles would require a double wave tensor. It represents a particle which has equal probability of having either spin.

It is perhaps rather surprising that there is no corresponding separation of phase space for positive and negative charges. The physical reason appears to be that the conception of phase space is bound up with the conception of statistical equilibrium, and there can be no equilibrium if the charges (or the probabilities of positive and negative charges) do not balance. Equivalently it is attributable to our result that electrically saturated space satisfies different reality conditions; and the phase space (if any) associated with it would require to be reinvestigated from the beginning. Although there is considerable analogy between opposite sign of charge and opposite spin, there is the fundamental distinction that opposite charges, unlike opposite spins, cannot be transformed into one another by relativity rotations.

CHAPTER VIII

THE DIFFERENTIAL WAVE EQUATION

8·1. Conservation of Probability.

In the terminology introduced by Dirac† a "state" of a system consists of a particular distribution of probability over the various possible configurations. The state is supposed to extend over all time. The probability distribution may vary with t, but always so that the integrated probability of all the configurations for a fixed value of t is unity. In other words probability is conserved.

We may treat the probability as a fluid occupying the configuration space, the probability of a given range of configurations being represented by the mass of fluid in a corresponding volume. The change of distribution of probability is then represented by a motion of the fluid.

The method of wave mechanics is to analyse the whole probability of the system, which must be unity, into the probabilities $p_a, p_b, p_c, \ldots$ of a set of elementary states $a, b, c, \ldots$. Then if $q_a(x_\mu, t)$ is the probability of the configuration x_μ at time t in the state a, the whole probability of the configuration x_μ at time t is

$$p_a . q_a(x_\mu, t) + p_b . q_b(x_\mu, t) + p_c . q_c(x_\mu, t) + \ldots. \qquad (8·11)$$

A *perturbation* of the system means a variation of $p_a, p_b, p_c, \ldots$, subject to their sum remaining unity.

The object of this device is to separate the mathematics of interaction from the mathematics of structure; for the influence of extraneous bodies is described by changes of the factors p, and the functions q which describe the structures of the various states remain unaffected.

Since the elementary states are introduced for analytical purposes, we may impose on them such limitations as appear advantageous. It would be possible to contemplate discontinuous flow of probability, whereby probability is created at one point and disappears at another point in the state, subject to the total probability remaining constant; but in current quantum theory the states are assumed to be *continuous*, that is to say the fluid moves subject to the equation of continuity.

We introduce a vector j_μ whose component in the time direction gives the density of the probability fluid and whose space components give the density of its flux. The conservation of probability is then secured by the equation of continuity of the fluid

$$\operatorname{div} j_\mu = 0. \qquad (8·12)$$

† *Quantum Mechanics*, 1st ed.

This assumes that the configurations are represented in a space with Euclidean metric. More generally the density of probability and of its flux are represented by a vector density $\mathfrak{j}_\mu$, and the equation of continuity is

$$\operatorname{div} \mathfrak{j}_\mu = 0, \tag{8·13}$$

the covariant divergence of a vector density being the same as its ordinary divergence.

For the general development of the theory it is necessary to use the phase space and strain vectors treated in Chapter VII; but in order to connect our treatment with Dirac's theory, we first consider a simple particle whose "configurations" are determined by its coordinates x, y, z in ordinary space. The probability distribution in state a is then a function $q_a(x, y, z, t)$.

Since we admit perturbations, we have to recognise an independent variable s extraneous to the state as an argument for the perturbations, so that (8·11) becomes

$$\Sigma_a p_a(s) . q_a(x, y, z, t). \tag{8·14}$$

The distinction between s and t is often ignored; but it is necessary to attend to it when we consider Lorentz transformations. We may apply a Lorentz transformation to x, y, z, t, and thus obtain a new but equivalent description of the probability distribution in the state; but the function $p_a(s)$ is a factor applying to the whole domain of x, y, z, and the time s which appears in it is not associated with any particular values of x, y, z; there is therefore no possibility of applying a Lorentz transformation to s, and it is invariant for the internal transformations of the state.

The perturbations are imposed on the system from without, and occur at times fixed according to an extraneous time reckoning. If we like, we may choose the time t within the state so as to conform to this reckoning; but we then lose the possibility of applying Lorentz transformations to the state. It is undesirable to do this—at any rate at the present stage—since the study of the Lorentz transformations, initiated by Dirac, has contributed greatly to our understanding of the theory. To preserve this advance and exhibit it in its proper relation to the more general developments, we have to distinguish a relative time t in the state and an invariant time s which serves as a link with other systems.†

If we wish to make a relativity transformation which includes both the perturbed and the perturbing system, we must treat them as one combined system and analyse the probability distribution into elementary states of the combined system. But then, for this combined system, there will be an invariant time s forming the argument of the perturbations of the combined system by systems extraneous to it. However many systems we combine there is always an invariant time left over for the purpose of representing

† In the nomenclature introduced later in this chapter, t is a geometrical coordinate and s a dynamical coordinate.

the perturbation effects of systems not yet included. It would be idle to consider a system without making provision for perturbation from outside; for in that case we could never acquire knowledge of it—unless indeed it were so extensive as to include the brain of an observer.

We note in particular that the time t is reversible, but the time s is irreversible. The axes in the state can be rotated so that $t \to -t$, and there is no absolute distinction between waves travelling forward and waves travelling backward in t. The latter are oddly said to have negative energy or mass according to the technical definitions in quantum theory. But the interaction with other systems depends on the invariant time s; thus the rotation of t and the substitution of "negative mass" for positive mass does not signify an observable change—it would not be a relativity transformation if it did. Similar considerations apply to more complex systems, until we reach a system which includes a human brain and is observed from within instead of from without. Extraneous time s is then no longer needed. But the inclusion of the observer in the system automatically prevents the relativity transformation from being pushed so far as to turn t into $-t$; at least it is unusual to include among the admissible systems of description (§ 1·1) those of an observer whose consciousness runs backward in the adopted time reckoning, and who endeavours to predict the past from his memories of the future.

By (8·14) the probability of a configuration x_μ is a function of the two times s, t. The same duality of time occurs, for example, in perturbation theory in celestial mechanics; at each moment s there is an osculating orbit of a planet which professes to give the position of the planet for all times t between $-\infty$ and $+\infty$. These positions are not actually realised. To obtain the realised positions we have to associate corresponding values of s and t. This means that in the present problem we have to lay down in the state a series of loci $f(x, y, z, t) = s$ which give a "representation" of s in the state. In particular cases it could be arranged that $f(x, y, z, t) = t = s$; but, as already stated, this definite choice of t requires that we forgo the application of Lorentz transformations.

The coordinate s, which is primarily an extraneous time-variable but has also a representation in the state, is a connecting link between the system which is being described and the rest of the universe. This ideal type of connection cannot be exactly realised in practice. The approximation lies in the assumption that, whatever the source of the perturbation, its argument s is represented by the same series of loci $f(x, y, z, t)$ in the state. Practically this means that the velocity of propagation of all perturbations is assumed to be infinite. If we require greater accuracy, the only course is to amalgamate the perturbed and perturbing system and investigate the states of the combined system.

To insist on this greater accuracy would be a counsel of despair. It would be to tell the physicist that, since the whole universe is interrelated, it is no use his attempting to study anything less than the whole universe. It is obvious that it is profitable to study portions of the universe as isolated systems. The method of states provides for this, and at the same time makes provision for re-attaching these systems to the rest of the universe in an approximate way, so that they may not be entirely cut off from observation.

It would be possible to choose the set of elementary states a, b, c, ... in many different ways. In practice therefore certain limitations are imposed. In particular the states are so chosen that the wave tensor $\Sigma E_\mu j_\mu$ corresponding to the probability vector is factorisable. The state is then said to be *pure*. Further, the whole set of states must be complete without being redundant so that, when the probability distribution of the configurations at time t is given, the coefficients p_a, p_b, ... are uniquely determined by it. The conditions of purity, completeness and non-redundancy require that the states shall correspond to a spectral set of operators. The spectral set is, of course, much more general than that which we have introduced for resolving an isolated wave tensor (§ 5·7). We shall not enter into the detailed treatment of these conditions, since our theory here coalesces with current quantum theory.

8·2. The Divergence Condition.

For a system describable by simple wave tensors the probability vector j_μ representing a pure state will be of the form

$$J = \psi\chi^* = \Sigma j_\mu E_\mu .$$

We introduce two divergence operators

$$\nabla = \sum_1^{16} E_\mu (\partial/\partial x_\mu), \quad \nabla^* = \sum_1^{16} (\delta/\delta x_\mu) E_\mu, \tag{8·21}$$

where $\delta/\delta x_\mu$ signifies $\partial/\partial x_\mu$ written *after its operand*. Then since $j_\mu = -\frac{1}{4}\chi^* E_\mu \psi$,

$$\frac{\partial j_\mu}{\partial x_\mu} = -\tfrac{1}{4}\chi^* \frac{\delta}{\delta x_\mu} E_\mu \psi - \tfrac{1}{4}\chi^* E_\mu \frac{\partial}{\partial x_\mu}\psi,$$

so that, summing for $\mu = 1, 2, \ldots, 16$,

$$\operatorname{div} j = \Sigma (\partial j_\mu/\partial x_\mu) = -\tfrac{1}{4}\chi^* (\nabla^* + \nabla)\psi. \tag{8·22}$$

It is understood that the same frame E_μ is used throughout the whole domain of x_μ in which ψ and χ^* extend.

The result of performing the operation ∇ on ψ is another four-valued quantity ω. We can always find a matrix M such that $\omega = M\psi$. Generally M will be a function of the coordinates. We therefore set

$$\nabla\psi = M\psi, \quad \chi^*\nabla^* = \chi^* M^*, \tag{8·23}$$

so that

$$\operatorname{div} j = -\tfrac{1}{4}\chi^* (M^* + M)\psi. \tag{8·24}$$

We ensure the vanishing of div$j$ by setting $M^* = -M$; so that (8·23) gives the differential wave equations

$$(\nabla - M)\psi = 0, \quad \chi^*(-\nabla^* - M) = 0. \tag{8·25}$$

These wave equations will be invariant for all wave-tensor transformations, if ∇, ∇^* and M are mixed wave tensors. This requires that $\partial/\partial x_\mu$ shall be a complete space vector; that is to say, when a relativity transformation is applied, it is transformed into a new array of operators $\partial/\partial x_\mu'$ which are linear functions of $\partial/\partial x_\mu$, just as if it were a numerical vector. The condition that $\partial/\partial x_\mu$ is a complete space vector is equivalent to the condition that x_μ is a complete space vector.

If we limit ourselves to solutions of (8·25) which are functions of four rectangular coordinates (x_1, x_2, x_3, x_4), the equations reduce to

$$\left.\begin{aligned} \left(E_1\frac{\partial}{\partial x_1} + E_2\frac{\partial}{\partial x_2} + E_3\frac{\partial}{\partial x_3} + E_4\frac{\partial}{\partial x_4} - M\right)\psi = 0, \\ \chi^*\left(-E_1\frac{\delta}{\delta x_1} - E_2\frac{\delta}{\delta x_2} - E_3\frac{\delta}{\delta x_3} - E_4\frac{\delta}{\delta x_4} - M\right) = 0. \end{aligned}\right\} \tag{8·261}$$

These equations are invariant only for the six relativity rotations in four dimensions; because the other relativity transformations (applicable to (8·25)) would reintroduce the terms that have been dropped.

Dirac's wave equations for an electron of proper mass m in an electromagnetic field, which gives it an energy and momentum κ_μ, can be written as

$$\left.\begin{aligned} \left\{\sum_1^4 E_\mu\left(-i\frac{\partial}{\partial x_\mu} + \kappa_\mu\right) - m\right\}\psi = 0, \\ \chi^*\left\{\sum_1^4\left(i\frac{\delta}{\delta x_\mu} + \kappa_\mu\right) - m\right\} = 0. \end{aligned}\right\} \tag{8·262}$$

Comparing (8·261) and (8·262), we obtain the identification

$$iM = -m + E_1\kappa_1 + E_2\kappa_2 + E_3\kappa_3 + E_4\kappa_4. \tag{8·271}$$

Writing as usual $M = \Sigma m_\mu E_\mu$, we have

$$m_{16} = m \qquad (m_1, m_2, m_3, m_4) = -i(\kappa_1, \kappa_2, \kappa_3, \kappa_4). \tag{8·272}$$

We have made this comparison with Dirac's equation in order to ascertain the current nomenclature for the components of our space vector M. The consequences of his equation have been worked out and compared with experiment, so that we know how the quantities m, κ_μ contained in it manifest themselves observationally. Equation (8·272) transfers this knowledge and the consequent nomenclature to our own equations (8·261). It would be foreign to our plan to intermingle the current semi-empirical theory with the purely deductive theory that we are developing, and our reference to the current equations is for identification purposes only. In current theory m is supposed to be a constant independent of the coordinates.

We have as yet shown no reason why m should be constant; that is a consequence of the dynamical equations, and will be proved in § 9·2; but it will be assumed here in anticipation.

If
$$p_\mu = -i\frac{\partial}{\partial x_\mu} + \kappa_\mu = i\frac{\delta}{\delta x_\mu} + \kappa_\mu, \tag{8·28}$$
(8·262) reduces to
$$(\Sigma E_\mu p_\mu - m)\psi = 0, \quad \chi^*(\Sigma E_\mu p_\mu - m) = 0,$$
which formally agrees with the simple wave equation as given in (5·13), (5·14). But this is only a real agreement if (8·28) gives an algebraic value of p_μ; for in (5·14) the p_μ are necessarily algebraic coefficients.

In order that the p_μ defined by (8·28) may be algebraic, ψ and χ^* must be eigensymbols of $\partial/\partial x_\mu$. This requires that ψ shall be a function of x_μ of the form
$$\psi(x_1, x_2, x_3, x_4) = f(x_1, x_2, x_3, x_4).\psi_0, \tag{8·285}$$
where f is an algebraic function and ψ_0 a constant wave vector. We shall call wave functions of the form (8·285) *algebraic wave functions*.

It would be undesirable to exclude non-algebraic wave functions. We shall find, for example, that the wave functions giving the steady states of a hydrogen atom are non-algebraic. Thus for unrestricted wave functions the momenta p_μ are non-algebraic; and the identification of p_μ with the algebraic coefficients j_μ in § 5·3 does not apply.

Thus in general we have two independent equations
$$\left(\sum_1^4 E_\mu j_\mu - m\right)\psi = 0, \tag{8·291}$$
$$\left(\sum_1^4 E_\mu p_\mu - m\right)\psi = 0. \tag{8·292}$$
The first is an identity, except that it is implied that the axes are so chosen that $j_5 = 0$. The second expresses the conservation of probability. For algebraic wave functions the two equations are the same; so that the momentum vector is the same as the stream vector except for a numerical factor. For non-algebraic wave functions the two equations are distinct; and no comparison between the stream vector and momentum vector is possible, since the components of the latter are non-algebraic quantities. We can rewrite (8·292), setting
$$\sum_1^4 E_\mu p_\mu = \sum_1^{16} E_\mu p_\mu',$$
where the p_μ' are algebraic coefficients; then the momentum vector can be regarded as a complete space vector p_μ', which has been formally reduced to four dimensions by the device of admitting symbolic components.

If there is no electromagnetic field, (8·262) has the elementary solutions
$$\psi = e^{i(p_1x_1+p_2x_2+p_3x_3-p_0x_0)}\psi_0, \quad \chi^* = e^{-i(p_1x_1+p_2x_2+p_3x_3-p_0x_0)}\chi_0^*, \tag{8·293}$$

where the p_μ are algebraic, and ψ_0, $\chi_0{}^*$ are solutions of the elementary wave equations (5·13).

In this book we shall use non-algebraic wave functions in special investigations, but in pursuing the more fundamental problems we shall generally limit ourselves to algebraic wave functions. That is because we are especially concerned with the borderland of relativity and quantum theory, and, as explained in the introduction, their meeting-point is to be found in the most uniform conditions.

According to (8·28) there are two different formulae for p_μ. At present the position (final or initial) of the wave vector is a sufficient indication which formula applies. But it is hampering to have a definition of p_μ which prevents us from changing the position of the wave vector; and we shall later describe the difference of the two operators in a more general way (§ 8·6).

8·3. Covariant Differentiation.

We do not propose to generalise our formulae to apply to all kinds of curvilinear coordinates; but there is one kind of curvilinear coordinate which it will be necessary to use, viz. an angular coordinate. Consider the transformation from rectangular coordinates (x_1, x_2) to polar coordinates (r, θ_{12}) in a plane; the question arises how this change is supposed to affect a general space vector j_μ and its wave-vector factors ψ, χ^*.

Our theory of ψ has been based on an orthogonal frame of reference, and it would be a grave complication to depart from an orthogonal frame. Therefore we do not consider the polar components of a vector in the sense of general relativity, but rather in the sense of elementary mechanics in which the "polar components" of a force are its rectangular components in the radial and transverse directions.† That is to say, our transformation will correspond to $(dx_1, dx_2) \to (dr, r\,d\theta_{12})$, not to $(dx_1, dx_2) \to (dr, d\theta_{12})$. Thus we retain an orthogonal frame, but the frame rotates as θ_{12} changes; and a local vector $\psi\chi^*$ is resolved orthogonally in continually changing directions.

We shall now find the covariant derivative of ψ with respect to θ_{12}. By (4·15) the transformation $\psi' = e^{\frac{1}{2}E_{12}\theta_{12}}\psi$ gives $t_1' = t_1 \cos\theta_{12} - t_2 \sin\theta_{12}$, which is the change of t_1 due to the axes having been rotated through an angle $-\theta_{12}$. In the present case, when the angular coordinate changes from 0 to θ_{12}, the axes are rotated in the same direction so that the corresponding transformation of ψ is

$$\psi' = e^{-\frac{1}{2}E_{12}\theta_{12}}\psi. \qquad (8·31)$$

This assumes that there has been no "real" change of ψ; that is to say, if we had kept to rectangular coordinates, ψ would have had the same value at

† There is a distinction between "a transformation of coordinates from (x, y) to (r, θ)" and "a transformation from rectangular coordinates (x, y) to polar coordinates (r, θ)". The first implies that all coordinates are to be treated alike as in general relativity; the second implies that they are to receive the distinctive treatment usually accorded to rectangular and polar coordinates. (Cf. *Mathematical Theory of Relativity*, § 16.)

(r, θ_{12}) as at $(r, 0)$. We call displacement without real change *parallel displacement*. Differentiating (8·31), the change due to parallel displacement is $\partial\psi'/\partial\theta_{12} = -\frac{1}{2}E_{12}\psi'$. For a general displacement we obtain the covariant derivative by subtracting from the apparent change $\partial\psi'/\partial\theta_{12}$ the change arising by parallel displacement. Thus the covariant derivative operator is

$$\left(\frac{\partial}{\partial\theta_{12}}\right)_c = \frac{\partial}{\partial\theta_{12}} + \tfrac{1}{2}E_{12}. \tag{8·321}$$

The operator $$M_{\mu\nu}' = -i\left(\frac{\partial}{\partial\theta_{\mu\nu}}\right)_c = -i\left(\frac{\partial}{\partial\theta_{\mu\nu}} + \tfrac{1}{2}E_{\mu\nu}\right) \tag{8·322}$$

is called the angular momentum conjugate to the coordinate $\theta_{\mu\nu}$. The term $-\frac{1}{2}iE_{\mu\nu}$ is called the *spin momentum*. It was originally discovered as a correction to the angular momentum, which had previously been assumed to be $-i\partial/\partial\theta_{\mu\nu}$. *We see that the spin momentum is merely the difference between the covariant derivative and the ordinary derivative.* It is a nominal addition to the angular momentum due to our non-relativistic outlook.

By (3·38) the expectation value of $E_{\mu\nu}$ is $ij_{\mu\nu}/j_{16}$, so that the expectation value of the spin momentum is $\frac{1}{2}j_{\mu\nu}/j_{16}$. The term "spin momentum" is primarily limited to the components j_{23}, j_{31}, j_{12}; but analytically all components are on the same footing, and we have a complete space vector $J/2j_{16}$ giving the part of the momentum (expectation value) which arises from the difference between covariant and ordinary derivatives. From this aspect the stream vector J is also a momentum vector. The identification in § 6·6 of its E_{23} component with mechanical spin is thus confirmed and elucidated.

For an initial contravariant wave vector χ^*, the corresponding operators are

$$\left(\frac{\delta}{\delta\theta_{\mu\nu}}\right)_c = \frac{\delta}{\delta\theta_{\mu\nu}} - \tfrac{1}{2}E_{\mu\nu}, \tag{8·331}$$

$$M_{\mu\nu}' = i\left(\frac{\delta}{\delta\theta_{\mu\nu}}\right)_c = i\left(\frac{\delta}{\delta\theta_{\mu\nu}} - \tfrac{1}{2}E_{\mu\nu}\right). \tag{8·332}$$

Hence the covariant derivative of a space vector $J = \psi\chi^*$ is

$$\left\{\left(\frac{\partial}{\partial\theta_{\mu\nu}} + \tfrac{1}{2}E_{\mu\nu}\right)\psi\right\}\chi^* + \psi\left\{\chi^*\left(\frac{\delta}{\delta\theta_{\mu\nu}} - \tfrac{1}{2}E_{\mu\nu}\right)\right\}$$
$$= \partial J/\partial\theta_{\mu\nu} + \tfrac{1}{2}(E_{\mu\nu}J - JE_{\mu\nu}). \tag{8·34}$$

The angular momenta generally referred to in quantum mechanics are conjugate not to the angles $\theta_{\mu\nu}$, but to angular parameters $-\alpha_{\mu\nu}$ introduced in the following way. Consider a distribution of ψ constituting an elementary state of a system. If the boundary conditions, extraneous electromagnetic fields, etc. are symmetrical in the plane of $\theta_{\mu\nu}$, we obtain another elementary state by rotating the whole distribution of ψ through an angle $\alpha_{\mu\nu}$ in that plane. Thus we obtain a series of distributions $\psi(x_1, x_2, x_3, x_4, \alpha_{\mu\nu})$, x_1, x_2,

x_3, x_4 being coordinates, and $\alpha_{\mu\nu}$ a parameter distinguishing one distribution from another. In place of x_1, x_2, x_3, x_4 we can use polar coordinates, including the angle $\theta_{\mu\nu}$. We have

$$\psi(x_\sigma, \alpha_{\mu\nu}) = \exp(\tfrac{1}{2}E_{\mu\nu}\alpha_{\mu\nu} - \alpha_{\mu\nu}\partial/\partial\theta_{\mu\nu}).\psi(x_\sigma, 0), \qquad (8{\cdot}35)$$

since the first term in the exponential gives the effect on ψ of rotating the axes backward through $\alpha_{\mu\nu}$, which is equivalent to rotating the whole distribution forward through $\alpha_{\mu\nu}$, and the second term is introduced because the point $\theta_{\mu\nu}$ in the new distribution corresponds to the point $\theta_{\mu\nu} - \alpha_{\mu\nu}$ in the old distribution.

Differentiating (8·35) with respect to $\alpha_{\mu\nu}$, we have

$$\frac{\partial}{\partial\alpha_{\mu\nu}} = -\left(\frac{\partial}{\partial\theta_{\mu\nu}} - \tfrac{1}{2}E_{\mu\nu}\right), \qquad (8{\cdot}361)$$

and the angular momentum conjugate to $-\alpha_{\mu\nu}$ is

$$M_{\mu\nu} = i\frac{\partial}{\partial\alpha_{\mu\nu}} = -i\left(\frac{\partial}{\partial\theta_{\mu\nu}} - \tfrac{1}{2}E_{\mu\nu}\right). \qquad (8{\cdot}362)$$

For χ^* the corresponding operator is

$$M_{\mu\nu} = i\left(\frac{\delta}{\delta\theta_{\mu\nu}} + \tfrac{1}{2}E_{\mu\nu}\right). \qquad (8{\cdot}363)$$

We shall call $\alpha_{\mu\nu}$ a *dynamical coordinate*, and distinguish $M_{\mu\nu}'$ and $M_{\mu\nu}$ as *geometrical* and *dynamical* angular momenta, respectively. The importance of $M_{\mu\nu}$ is that (in the symmetrical conditions already postulated) it commutes with the hamiltonian; this, as we shall see later, makes it a constant of the motion of the system, which is then said to possess an integral of angular momentum. But when the conditions are not symmetrical, and no integral of angular momentum exists, there is no reason to suppose that (8·362) represents angular momentum of any kind. It is therefore rather misleading to say that $-i(\partial/\partial\theta_{\mu\nu} - \frac{1}{2}E_{\mu\nu})$ is the angular momentum of a system; it is a form to which the dynamical momentum reduces in particular cases when it happens to be constant. On the other hand, $M_{\mu\nu}'$ has the same interpretation for symmetrical and for unsymmetrical systems. Whilst it has apparently little connection with momentum as conceived in classical mechanics, it is a natural generalisation of the quantum theory definition.

To reach the dynamical outlook we must promote α_{12}, originally introduced as a parameter, to be a coordinate; so that ψ is a function of five coordinates $\alpha_{12}, x_1, x_2, x_3, x_4$, or in polar coordinates $\alpha_{12}, \theta_{12}, \phi, r, t$. (The configurations occupy four dimensions θ_{12}, ϕ, r, t, as before.) This means that a series of elementary four-dimensional states is run together to form a single five-dimensional state; and we adopt a new dissection into states

in which the five-dimensional state is regarded as elementary.† In combining the four-dimensional states we attribute to them a uniform probability distribution in α_{12}. We can now assign only one probability factor p to the whole five-dimensional state, whereas formerly we could assign different probability factors p_α to each orientation α of the four-dimensional state. This sacrifice corresponds to the fact that, the conditions being perfectly symmetrical, it is impossible to distinguish the orientation observationally, and therefore we never have occasion to consider a modification of the initially assumed uniform probability distribution in orientation ("*a priori* probability") through additional information furnished by observation.

The usual method of obtaining (8·362) is to consider a form of hamiltonian, which by its symmetry ensures that the dynamical angular momentum is constant, and then to identify an analytical expression which turns out to be constant as the angular momentum. This gives no indication of the expression for a non-constant angular momentum. For example, in an asymmetric electromagnetic field, systems, besides being deformed by the field, will tend to orient themselves in a certain way. Even if we ignore the deformation and assume that exactly similar four-dimensional states can exist in different orientations α, it will be necessary to insert in (8·35) a probability factor $p_\alpha{}^{\frac{1}{2}}$ representing the unequal probability of distribution of the different orientations; so that there will be an additional term in $\partial\psi/\partial\alpha$. It is clear therefore that (8·362) is not the correct expression for the angular momentum in unsymmetrical conditions.

In setting $\operatorname{div} j = \Sigma\,\partial j_\mu/\partial x_\mu$ (§ 8·2), we assumed that the x_μ are rectangular coordinates, the volume element being taken to be $dV = dx_1 dx_2 dx_3 \ldots$ so that the probability or probability flux belonging to an element is $j_\mu dV/dx_\mu$. If angular coordinates are used, two courses are open. The simplest is to introduce the vector density $\mathfrak{j}_\mu = j_\mu\sqrt{-g}$; the equation of conservation is then $\Sigma\,(\partial \mathfrak{j}_\mu/\partial x_\mu) = 0$. Otherwise we must substitute covariant derivatives in place of ordinary derivatives in ∇.

It is instructive to check the agreement of the two methods. If E_r is the matrix corresponding to the radial direction, and E_θ the matrix for a rotation in a plane containing the radius (therefore anticommuting with E_r), the matrix for the corresponding transverse direction is $E_\theta E_r$. The corresponding term in ∇ is

$$E_\theta E_r\left(\frac{\partial}{r\partial\theta}\right)_c = \frac{E_\theta E_r}{r}\left(\frac{\partial}{\partial\theta} + \tfrac{1}{2}E_\theta\right) \quad \text{by (8·321)}$$

$$= E_\theta E_r \frac{\partial}{r\partial\theta} + \frac{E_r}{2r}. \qquad (8\cdot37)$$

† The running together of the configurations along a singular line in § 7·9 to form a single "state of spin" of an elementary particle may be regarded as an elementary example of this procedure.

Thus for each rectangular coordinate replaced by an angular variable, $E_r/2r$ is added to ∇. If there are n angular variables, the result is to change the radial term $E_r\partial/\partial r$ to

$$E_r\left(\frac{\partial}{\partial r}+\frac{\frac{1}{2}n}{r}\right). \qquad (8{\cdot}38)$$

If we set $\psi=r^{-\frac{1}{2}n}\phi$ and take ϕ as a new wave function, the extra term $\frac{1}{2}n/r$ is eliminated. We find that χ^* behaves in the same way; so that, setting $\chi^*=r^{-\frac{1}{2}n}\omega^*$, the extra term is eliminated in the equation for ω^*. (In treating χ^* we must notice that $E_\theta E_r$ is replaced by $E_r E_\theta$, and this cancels the change of sign of the spin correction.) Accordingly we have a new combination $\mathfrak{J}=\phi\omega^*=r^nJ$ which satisfies the simple divergence equation without the added terms. This will agree with our former result if $\mathfrak{J}$ is the vector density $J\sqrt{-g}$; hence $\sqrt{-g}$ should be equal to r^n. This is correct, because each substitution $dx_\mu=r\,d\theta_\mu$ contributes a factor r to $\sqrt{-g}$.

8·4. General Dynamical Equations.

Since our symbolic calculus has been extended by the introduction of differential operators $\partial/\partial x_\mu$, which I will call D-symbols, it is necessary to refer again to the formal definitions in § 2·1.

A symbol which commutes with every symbol in the calculus will be called an algebraic number as heretofore. A symbol which commutes with all symbols other than D-symbols will be called an *algebraic function*.†

Since $D(\psi\chi)\neq(D\psi)\chi$, the D-symbols do not obey the associative law of multiplication. If $D_x=\partial/\partial x$, $D_x f\chi=\partial(f\chi)/\partial x$. Hence

$$(\partial f/\partial x)\chi=(D_x f-fD_x)\chi$$

whatever χ may be. We have therefore in all cases

$$\partial f/\partial x=D_x f-fD_x. \qquad (8{\cdot}41)$$

The introduction of D-symbols leads us to contemplate a wider variety of tensor transformations, i.e. a wider variety of systems of description of a physical system. We have been using transformations of the form $q=e^\Theta$, where Θ is a matrix or more generally an E-number. It is natural now to admit still more general transformations in which Θ may be any combination of symbols, including D-symbols. In particular we consider the transformation

$$q=e^{iWs}, \qquad (8{\cdot}421)$$

where W is any symbolic expression and s is an algebraic parameter.

Let χ^*, ψ be initial and final eigensymbols of W, the eigenvalue m being the same for both, so that

$$\chi^*(W-m)=0, \quad (W-m)\psi=0. \qquad (8{\cdot}422)$$

† The term "algebraic function" does not include "algebraic wave functions", defined in (8·285) as the product of an algebraic function and a constant wave vector.

Transforming these as contravariant and covariant wave vectors respectively, we have

$$\left.\begin{aligned}\chi^{*\prime} &= \chi^* e^{-iWs} = \chi^* e^{-ims},\\ \psi' &= e^{iWs}\psi = e^{ims}\psi,\end{aligned}\right\} \tag{8·423}$$

so that, if $J = \psi\chi^*$, $J' = J$.

If T is any other mixed tensor of the same class as J,

$$T' = e^{iWs} T e^{-iWs}. \tag{8·424}$$

Hence for an infinitesimal change ds

$$T + dT = (1 + iW\,ds)\,T\,(1 - iW\,ds),$$

so that

$$dT/ds = i(WT - TW). \tag{8·43}$$

To apply this, we regard s as a coordinate, and consider a physical system S described by mixed wave tensors T or the equivalent space vectors. When (8·43) is satisfied, displacement along s is *parallel displacement* of the system; that is to say, the tensors describing the system are unchanged except that they undergo a common transformation which we interpret as a transformation of reference frame.

It is important to understand the significance of this association of a tensor transformation with every displacement. If the system S is the only system contemplated, a change of the system of description has no useful purpose; but then it is idle to talk of displacing the system, since there is nothing to which the displacement can be referred. It is therefore presupposed that there exists besides S a reference system S' to enable the displacement to be recognised; and the displacement constitutes an intrinsic alteration or strain of the combined system S, S'. The transformation of the system of description, applied to S but not to S', expresses the fact that although S is intrinsically unchanged, its relation to S' has been altered.

In general relativity we are familiar with this change of the system of description which necessarily accompanies every observable displacement. It is expressed by the fact that an extended system of rectangular coordinates is impossible, or equivalently that the curvature of physical space cannot vanish. The effect of curvature is that displacements in different directions do not commute. In symbolic calculus we are indifferent to geometrical pictures, and express the same thing more directly by associating non-commuting operators with the displacements. The change of description is directly associated with non-commutation of operators, as may be seen from (8·43); if the operator W associated with the displacement s commutes with all the tensors T of the physical system no change of description occurs.

The displacements whose associated operators are E-numbers can be represented in Riemannian space. It is, I think, improbable that the

displacements associated with the more general operators now admitted are representable in Riemannian space. If we wish to represent them graphically we must adopt whatever form of geometry is necessary to provide representation of their non-commutative relations. In practice, however, we are not much concerned with the totality of transformations of the type (8·421); we have only to pick out a few special forms which yield comparatively simple systems.†

The coordinate s is a dynamical coordinate of the same type as the coordinate $\alpha_{\mu\nu}$ introduced in § 8·3. It will be remembered that we there considered a probability distribution in space-time and changed its orientation through an angle $\alpha_{\mu\nu}$. Since the distribution was intrinsically unaltered, this was a parallel displacement. The space vectors describing the probability distribution at *all* points of the system underwent the same transformation (8·35). Since for any wave vectors ψ, χ^*, the infinitesimal transformation is

$$\psi + d\psi = (1 + iW ds)\psi, \quad \chi^* + d\chi^* = \chi^*(1 - iW ds),$$

we have

$$W = -i\partial/\partial s = i\delta/\delta s, \tag{8·44}$$

so that W is the momentum conjugate to s. It will be seen that $M_{\mu\nu}$ and $\alpha_{\mu\nu}$ form a particular case of W and s.

The distinction between geometrical and dynamical coordinates is rather obscured by the fact that in the most familiar system of coordinates, viz. rectangular coordinates, the conjugate momenta are the same. The distinction is necessary in angular and other curvilinear coordinates because the conjugate momenta differ, and indeed are scarcely comparable in conception. Further, by generalising our operators, we have introduced dynamical coordinates which may not be representable in the same space as the geometrical coordinates and may therefore have no counterpart in Riemannian geometry. The essential difference is that the geometrical coordinates express the internal relations of a system or probability distribution, and the dynamical coordinates express its relations to external objects. Geometrical coordinates are internal; dynamical coordinates are external. In elementary theory the only changes of external relations contemplated are those corresponding to change of position or orientation; but by means of general dynamical coordinates we can introduce external variables which more closely correspond to the internal strain vectors of the extended system which comprises the external reference objects.

Equation (8·43) is a point of junction of the present theory with Dirac's theory. He arrived at it by seeking what he regarded as the most natural adaptation of the classical equations of motion to quantum conditions.

† E.g., the forms W, U_1, U_2, U_3 in § 9·2.

8·5. Extension to Four Dynamical Coordinates.

Let W, U_1, U_2, U_3 be independent symbolic expressions which mutually commute. We may expect to be able to find a common eigensymbol for them. It was shown rigorously in § 3·7 (*e*) that a common eigensymbol can be found if the commuting symbols are matrices, and the same proof applies if they are symbols which satisfy algebraic equations. It is not clear that it applies to operators such as $\partial/\partial x$ which have an infinite number of eigenvalues. But no inconsistency can arise through postulating a common eigensymbol for mutually commuting symbols, and the only limitation on the invention of symbols is that we must not ascribe to them properties which are not self-consistent. It would seem therefore that, if we cannot find a common eigensymbol for W, U_1, U_2, U_3, we are at liberty to invent one—just as we have invented a square root of -1 in algebra.

For us it is sufficient that there exist important applications in which initial and final eigensymbols ψ, χ^* of the four operators can be found. An example will be given in § 9·2. Let the eigenvalues be m, μ_1, μ_2, μ_3. Then the transformation (8·423) can be extended to

$$\psi' = e^{i(Ws+U_1 s_1+U_2 s_2+U_3 s_3)}\psi = e^{i(ms+\mu_1 s_1+\mu_2 s_2+\mu_3 s_3)}\psi. \qquad (8·51)$$

As before the product $J=\psi\chi^*$ is unaltered by the transformation. For other mixed tensors T we have in place of (8·43)

$$\partial T/\partial s = i(WT-TW), \quad \partial T/\partial s_1 = i(U_1 T - TU_1), \text{ etc.} \qquad (8·52)$$

and W, U_1, U_2, U_3 are the momenta conjugate to the dynamical coordinates s, s_1, s_2, s_3. By (8·52) they are constant over the whole domain of (s, s_1, s_2, s_3).

We shall call the four-dimensional domain of the coordinates s, s_1, s_2, s_3 an S-space. It is a very simple kind of space—perhaps simpler than Euclidean space—but it is unfamiliar since the axes in it are antiperpendicular. All displacements in it commute, so that it is pictured as flat; but in other respects it is not comparable with the space of ordinary conception.

An elementary example of an S-space is obtained by taking

$$W, U_1, U_2, U_3 = (E_{16}, E_\alpha, E_\beta, E_\gamma)\, m,$$

where E_α, E_β, E_γ form an anti-triad. We have then for all wave vectors

$$\psi(s, s_\alpha, s_\beta, s_\gamma) = e^{(E_{16}s+E_\alpha s_\alpha+E_\beta s_\beta+E_\gamma s_\gamma)m}\psi_0, \qquad (8·53)$$

which reduces to $$\psi(s, s_\alpha, s_\beta, s_\gamma) = e^{im(s\pm s_\alpha\pm s_\beta\pm s_\gamma)}\psi_0, \qquad (8·54)$$

when ψ_0 is a common eigensymbol of E_α, E_β, E_γ.

Alternatively we can describe the same domain by *spectral coordinates* s_a, s_b, s_c, s_d conjugate to momenta J_a, J_b, J_c, J_d defined as in (5·71).

In the most elementary problems we cannot have more than four dynamical coordinates, since not more than four independent E-numbers can mutually commute. For that reason we have chosen to consider four

dynamical coordinates in this section. The same general dynamical theory applies to any number of coordinates; but the problems involving four coordinates will be classed as one-body problems, and therefore come first in our order of treatment.

Of the four dynamical coordinates one is singled out to be the proper time s, and its conjugate momentum W is called the hamiltonian.† The reason for this selection must lie outside the system itself; for the dynamical equations (8·52) are perfectly symmetrical. The distinctive property of s can only appear when we contemplate the system in relation to other systems. It is, as we have seen (§ 8·1), the argument of the perturbations of and by other systems. The peculiarity that the system "goes on" in s, whereas it is merely extended in s_1, s_2, s_3, is explained if it is through s that changes in the system are linked to changes in the external world and therefore ultimately to the time sequence in consciousness.

The principle that the separate physical systems into which we dissect the universe shall each have just one coordinate in common with the rest, is valuable as expressing the conceptions which are the basis of our nomenclature. It is not so important that it should be fulfilled rigorously, since any supplementary coupling can be dealt with by perturbation methods. In adopting s as the unique link we assume an idealised standard environment of the system, such that any change in the environment produces effects which occur simultaneously in all parts of the system according to the time reckoning s. In special cases we may have to treat an environment which deviates markedly from this standard. If light waves are falling on an atom in a particular direction, we should take account of the fact that the perturbation travels across the atom with the velocity of light. It would seem therefore that the distinction between s, s_1, s_2, s_3 is a matter of degree; and that all four coordinates afford potential linkages with external systems of appropriate character, though only one is called into play by the standard environment which our equations presuppose.

To sum up: it is idle to treat in our equations a system supposed to have no interaction with its environment, since the interaction is the only thing about the system which concerns observational physics. On the other hand, it is not necessary to go to the other extreme and treat a system with an environment of the most general kind that can occur in nature. Just as we begin by studying the simplest systems, so we begin by studying the simplest form of environment, capable of introducing only the simplest type of perturbation. Under these conditions one coordinate s plays a unique rôle, and becomes distinguished from s_1, s_2, s_3.

† In our nomenclature. On all points which concern the relations of s and t our outlook differs so much from the current theory that comparison of nomenclature is scarcely possible.

8·6. The Differential Wave Equation for a Strain Vector.

The tetrad of matrices in the wave equations (8·261) or (8·262) may be taken to be either E_1, E_2, E_3, E_4 or $E_{15}, E_{25}, E_{35}, E_{45}$. The choice is not entirely a matter of indifference, because the two tetrads lead to different reality conditions. When the p_μ are algebraic, the former is the appropriate tetrad for a (four-dimensional) vector density $\psi\chi^*$, and the latter for a space vector $\psi\chi^*$, by (6·51). From considerations of continuity the same distinction must hold for non-algebraic wave functions.

Accordingly the wave equations without electromagnetic field for a space vector $\psi\chi^*$ are

$$\left(E_{15}\frac{\partial}{\partial x_1}+E_{25}\frac{\partial}{\partial x_2}+E_{35}\frac{\partial}{\partial x_3}+E_{45}\frac{\partial}{\partial x_4}-im\right)\psi=0, \qquad (8\cdot611)$$

$$\chi^*\left(E_{15}\frac{\delta}{\delta x_1}+E_{25}\frac{\delta}{\delta x_2}+E_{35}\frac{\delta}{\delta x_3}+E_{45}\frac{\delta}{\delta x_4}+im\right)=0. \qquad (8\cdot612)$$

Let $\phi^*=i\chi^*E_{45}$, so that $S=\psi\phi^*$ is the associated strain vector. Substituting in (8·612) we obtain

$$\phi^*\left(E_{14}\frac{\delta}{\delta x_1}+E_{24}\frac{\delta}{\delta x_2}+E_{34}\frac{\delta}{\delta x_3}+\frac{\delta}{\delta x_4}-iE_{45}m\right)=0, \qquad (8\cdot62)$$

which is equivalent to

$$\left(E_{14}\frac{\partial}{\partial x_1}+E_{24}\frac{\partial}{\partial x_2}+E_{34}\frac{\partial}{\partial x_3}+\frac{\partial}{\partial x_4}+iE_{45}m\right)\phi=0, \qquad (8\cdot631)$$

since E_{45} is the only antisymmetrical matrix in (8·62). Multiplying (8·611) initially by $-E_{45}$ we obtain

$$\left(E_{14}\frac{\partial}{\partial x_1}+E_{24}\frac{\partial}{\partial x_2}+E_{34}\frac{\partial}{\partial x_3}+\frac{\partial}{\partial x_4}+iE_{45}m\right)\psi=0. \qquad (8\cdot632)$$

Thus the covariant wave vectors ψ, ϕ are solutions of the *same* differential equation.

The equations for ψ and ϕ are, however, not the same when there is an electromagnetic field. In (8·611) and (8·612), m is replaced by

$$m-E_{15}\kappa_1-E_{25}\kappa_2-E_{35}\kappa_3-E_{45}\kappa_4;$$

so that in (8·62) and (8·632), $E_{45}m$ is replaced by

$$E_{45}m+E_{14}\kappa_1+E_{24}\kappa_2+E_{34}\kappa_3+\kappa_4.$$

But in passing from (8·62) to (8·631) the reversal of sign applies only to $E_{45}m$. The equations for ψ and ϕ, including electromagnetic terms, are most conveniently written

$$\left\{E_{14}\left(-i\frac{\partial}{\partial x_1}+\kappa_1\right)+\ldots+\left(-i\frac{\partial}{\partial x_4}+\kappa_4\right)+E_{45}m\right\}\psi=0, \qquad (8\cdot641)$$

$$\left\{E_{14}\left(i\frac{\partial}{\partial x_1}+\kappa_1\right)+\ldots+\left(i\frac{\partial}{\partial x_4}+\kappa_4\right)-E_{45}m\right\}\phi=0. \qquad (8\cdot642)$$

Remembering that i is contained implicitly in x_4, κ_4 and in the imaginary matrices E_{14}, E_{24}, E_{34}, we see that if the sign of (8·642) is reversed it becomes the complex conjugate of (8·641). So that if $\psi(x_1, x_2, x_3, t)$ is a solution of (8·641), its complex conjugate is a solution of (8·642). There are, of course, other solutions of (8·642), and it is not necessary to suppose that the two wave vectors representing a particle are complex conjugates.

The wave functions adopted by Dirac† and used, I think, in all current treatises are ψ, ϕ, not ψ, χ. Even when $\kappa_\mu = 0$, so that ψ and ϕ satisfy the same equation (8·631), they are taken to be different solutions representing waves travelling in opposite directions in four dimensions. This comes about because the momentum operators applying to them have been defined differently in (8·28); so that when the momentum has a given value (the same for ψ and ϕ) different functions are required.

We shall call wave functions whose momentum operator is $-i\partial/\partial x_\mu + \kappa_\mu$ *wave vectors of index* 1, and those whose momentum operator is $i\partial/\partial x_\mu + \kappa_\mu$ *wave vectors of index* -1. The former satisfy (8·641) and the latter (8·642). The definition will later be extended, so that a wave tensor is said to be of index n if its momentum operator is

$$p_\mu = -\frac{i}{n}\frac{\partial}{\partial x_\mu} + \kappa_\mu. \tag{8·65}$$

This applies to covariant, contravariant, initial or final wave tensors, $\partial/\partial x_\mu$ being changed to $\delta/\delta x_\mu$ when the tensor is written initially.

We have seen that ϕ *may be* the complex conjugate of ψ. Dirac goes further and defines ϕ as the complex conjugate of ψ. In the present theory there is no reason to impose this restriction, which is presumably a survival of the Hermitic conditions employed in Schrödinger's elementary theory. These are superseded by the reality conditions found in Chapter VI. For algebraic wave functions (8·285), ψ_0 and ϕ_0 can be chosen independently from the infinitude of solutions of the elementary wave equation, and there is in general a similar independence of non-algebraic wave functions.

Let us, however, consider for a moment the current theory which takes ϕ to be the complex conjugate of ψ. The full specification of the system is then contained in a single wave function ψ; for we do not add anything to the specification by inventing a special symbol for the complex conjugate. The system might equally well be specified by ϕ; then ψ is merely a symbol for the complex conjugate of ϕ. But ψ and ϕ represent waves travelling in opposite directions in four dimensions. It may be asked, Which are the real waves, or are there waves in both directions? The answer is that there are no *real* waves. I suppose that no one nowadays attributes objective existence to the waves described in wave mechanics.

† *Quantum Mechanics*, 2nd ed., p. 255, equations (9) and (10).

Since then the system is specified by a single wave vector ψ, the most natural wave tensor of the second rank furnished by it would seem to be $\psi\psi^*$ (or, with suffixes, $\psi_\alpha\psi_\beta$), i.e. the outer square of ψ. If

$$(-i\partial/\partial x_\mu+\kappa_\mu)\psi=p_\mu\psi,$$

where p_μ is algebraic,† we have

$$(-\tfrac{1}{2}i\partial/\partial x_\mu+\kappa_\mu)\psi\psi^*=p_\mu\psi\psi^*. \tag{8·66}$$

Thus $\psi\psi^*$ is of index 2. By introducing the complex conjugate an alternative wave tensor $\psi\phi^*$ of index 0 is obtained. Attention seems to have been devoted exclusively to the latter. It is of considerable importance; but we must not let it unduly divert attention from the primary wave tensor $\psi\psi^*$.

As already stated, we do not accept the limitation in Dirac's theory which reduces the specification of a particle to a single wave-vector function. The space vectors and strain vectors which comprise the ordinary vectors of physics are wave tensors of the second rank. We resolve these into pure constituents, which are factorisable into wave vectors. In general there is no reason to expect or require that the two factors shall be equal. I have, of course, no objection to the employment in quantum physics of wave tensors which are perfect squares, if these are appropriate to the problems which are studied—as is sometimes the case. But to regard it as more than a casual adaptation creates an artificial gulf between quantum theory and relativity theory, since there is no such limitation in the latter.

Thus a pure strain vector of index 2 will normally be the product of unequal factors $S_2=\psi_1\psi_2^*$. Exchanging one of these for its complex conjugate, we obtain an associated strain vector of index 0, $S_0=\psi_1\phi_2^*$. The latter is the strain vector we have been studying, but we shall now turn attention to S_2. S_2 has the advantage that there is no need to factorise it. The momentum is given by the operator (8·65), used as in (8·66). Factorisation is only needed for wave tensors of index 0. For them the operator (8·65) is indeterminate, and it is necessary to find a factor which is not of index 0 so as to obtain the momentum.

At present we treat only wave vectors of index ± 1. We may note, however, that there is a possibility of extending the theory to wave vectors of any index n, integral or fractional, p_μ being always given by (8·65).

8·7. Application to Phase Space.

In Chapter VII we have described a system by a strain vector which specifies simultaneously the configuration, the probability of the configuration, and the time. This cannot be the strain vector $S_0=\psi\phi^*$, which was introduced for the purpose of specifying the probability only, and is in fact independent

† Or, more generally, if p_μ is an E-number containing only space-like matrices and therefore commuting with ψ.

of the time. The strain vector which generates phase space is the associated strain vector of index 2, viz. $S_2 = \psi_1 \psi_2^*$.

Consider a free particle in field-free space, so that the ordinary wave functions are

$$\psi = e^{ims}\psi_0, \quad \phi = e^{-ims}\phi_0, \tag{8·711}$$

where

$$ms = p_1 x_1 + p_2 x_2 + p_3 x_3 - p_0 t.$$

Their product is the strain vector S_0. Alternatively we specify the particle by wave functions ψ_1, ψ_2, both of index 1,

$$\psi_1 = e^{ims}(\psi_1)_0, \quad \psi_2 = e^{ims}(\psi_2)_0. \tag{8·712}$$

Their product is the strain vector

$$S_2 = e^{2ims}(S_2)_0. \tag{8·72}$$

For a displacement ds, the only change produced in S_2 is a change of algebraic phase $d\theta_{16} = 2m\,ds$. If displacement in time (proper time) is expressed in the same linear measure in phase space as in ordinary space, so that $d\theta_{16} = ds/R$, we have $2m\,ds = ds/R$, so that

$$m = 1/2R, \tag{8·73}$$

where R is the radius of the phase space.

At first sight it is anomalous that the general displacement (dx_1, dx_2, dx_3, dx_4) should be interpreted as change of time only, and not change of configuration. But the plane wave solutions (8·712) presuppose flat space-time. If they are used in curved physical space-time, they must be restricted to regions not too large to be treated as flat. This means that the region of phase space which they cover is not too large for the distinction between the different configurations to be neglected. The apparent discrepancy is thus due to the nature of the approximation assumed in plane wave solutions; although they formally distinguish configurations by coordinates, they suppose that the distinctions when expressed by matrices are so inconsiderable that they can be neglected.

We have suggested (§ 5·4) that wave vectors are introduced mainly to secure purity of the wave tensors, and that many if not all of the problems of quantum theory could be solved by using the wave tensors directly. Current theory gives a rather fictitious importance to the vector factors, because it recognises only wave tensors of index 0. These do not contain the factor e^{ims}; and, since the energy m is one of the most important characteristics of a system, it is necessary to examine the factors in order to find it. This is avoided by the use of wave tensors of index 2, which contain the factor e^{2ims}. There is then no need to have recourse to the wave vectors, at any rate so far as the calculation of m is concerned.

The position may be summarised as follows. Certain properties of a system are naturally described by constant symbols, e.g. a steady distribution of probability or probability flux. When these are factorised, they are resolved

into components ψ, ϕ whose time factors cancel one another. But, since the time factors cannot be supposed to exist solely for the purpose of cancelling one another, this is a tacit admission that the constant symbols do not comprise the whole data of the system, and that in a complete description time factors must appear. Chapter VII gives this more compendious description in terms of strain vectors variable with the time; they are factorised into components whose time factors reinforce one another.

The general dynamical theory of a system described by strain vectors of index 2 is analogous to that developed for space vectors in §§ 8·4, 8·5. For a transformation $q=e^{iWs}$, we have

$$S'=\psi'\phi^{*\prime}=e^{iWds}.\psi\phi^{*}e^{i\overline{W}ds}=e^{iWds}Se^{i\overline{W}ds}, \tag{8·74}$$

where

$$W\phi=\phi^{*}\overline{W}. \tag{8·75}$$

The determination of $\overline{W}$ from W presents no difficulty, remembering that $(\partial/\partial x)\phi=\phi^{*}(\delta/\delta x)$. From (8·74) we obtain the general dynamical equation for strain vectors

$$dS/ds=i(WS+S\overline{W}). \tag{8·76}$$

The transformation introduces a strain common to all the strain vectors of the system. It may be regarded as defining *parallel strain* in the same way that the transformation of the space vectors in § 8·4 defines parallel displacement. The dynamical coordinate s measures a progressive parallel strain of the system. In elementary examples parallel strain is merely the internal aspect of what is externally regarded as parallel displacement of part of a system.

8·8. The Electromagnetic Potentials.

There exists an important transformation which leaves invariant the momentum vector $p_\mu=-i\partial/\partial x_\mu+\kappa_\mu$. Let

$$\psi'=e^{\frac{1}{2}E_{16}\theta_{16}}\psi=\lambda\psi, \tag{8·81}$$

where θ_{16} is an algebraic function of the coordinates x_μ. Then if

$$(-i\partial/\partial x_\mu+\kappa_\mu)\psi=p_\mu\psi, \quad (-i\partial/\partial x_\mu+\kappa_\mu')\psi'=p_\mu\psi',$$

we have† $$(-i\partial/\partial x_\mu+\kappa_\mu')\lambda\psi=p_\mu\lambda\psi=\lambda p_\mu\psi$$
$$=\lambda(-i\partial/\partial x_\mu+\kappa_\mu)\psi.$$

Hence

$$\kappa_\mu'=\kappa_\mu+i\partial(\log\lambda)/\partial x_\mu=\kappa_\mu-\tfrac{1}{2}\partial\theta_{16}/\partial x_\mu. \tag{8·82}$$

Accordingly the transformation $\psi\to\psi'$, $\kappa_\mu\to\kappa_\mu'$, defined by (8·81) and (8·82), leaves p_μ unaltered. The electromagnetic field of force is also unaltered, because the addition of an arbitrary gradient to the potentials has no effect on the force.

† It is understood that p_μ here denotes the *value* (number or matrix) of the component of the momentum vector, not the operational expression $-i\partial/\partial x_\mu+\kappa_\mu$. Thus p_μ commutes with λ.

The corresponding transformation of other wave tensors depends on the index n of the momentum operator $(-i/n)\,\partial/\partial x_\mu + \kappa_\mu$. The general law is

$$\psi' = \lambda^n \psi, \tag{8·83}$$

where n is the index of ψ (§ 8·6). Since the transformation is algebraic this applies to wave tensors of any rank ($S' = \lambda^n S$), the index of the tensor being the sum of the indices of its factors. In particular a wave tensor of index 0 is invariant.

The foregoing transformation will be called a *gauge transformation*, because it is the adaptation to wave mechanics of Weyl's gauge transformation in relativity theory. We may regard the electromagnetic potential κ_μ as having been created by a non-integrable gauge transformation of neutral space-time. If in (8·82) we determine θ_{16} so that $\partial\theta_{16}/\partial x_\mu = 2\kappa_\mu$, we have $\kappa_\mu' = 0$; that is to say, the gauge transformation removes the electromagnetic field—which may therefore be created by the inverse transformation. But the equations $\partial\theta_{16}/\partial x_\mu = 2\kappa_\mu$, determining the transformation, are non-integrable unless $\operatorname{curl} \kappa_\mu = 0$.

To justify the name "gauge transformation" we proceed as follows. If θ_{16} is imaginary, λ is real, and the strain vector S_2 which generates phase space is multiplied by a real algebraic factor λ^2. By § 7·7 this represents a change of the probability of a range of configurations at the point considered.† But the probability is also given by $\phi^*\psi dV$, where ψ, ϕ^* are the ordinary wave functions of indices 1 and -1.‡ Usually changes of probability are expressed by changes of the modifying factor $\phi^*\psi$; but in this transformation $\phi^*\psi$ is invariant, since it is of index 0. The transformation therefore changes the measure of volume to $dV' = \lambda^2 dV$. That is what is meant by a gauge transformation—a change of measure of volume (implying a change of the standard of length) without alteration of the coordinates. In terms of coordinates $dV = \sqrt{-g}\,.d\tau$, $dV' = \sqrt{-g'}\,.d\tau$, so that the transformation can also be expressed as $\sqrt{-g'} = \lambda^2\sqrt{-g}$.

In Weyl's theory it was taken for granted that changes of electromagnetic potential correspond to real changes of gauge. Wave mechanics introduces an important amendment. By (8·82) real changes of the electromagnetic potential (κ_1, κ_2, κ_3, κ_0) correspond to real values of θ_{16}, and hence to *imaginary changes of gauge*. The need for this amendment became obvious as soon as it was discovered in quantum theory that the significant combination is $-i\partial/\partial x_\mu + \kappa_\mu$, not $\partial/\partial x_\mu + \kappa_\mu$.

This amendment removes the only difficulties noticed in the unified

† Note that this interpretation only holds if λ is real. If λ is complex (θ_{16} real) the change affects the time coordinate and has no effect on the probability.

‡ For simplicity we suppose that the transformation redistributes the probability without altering the total amount, so that it is not necessary to re-normalise after the transformation.

gravitational-electromagnetic theory.† In the attempts to find a geometrical invariant representing the Action (*loc. cit.*, pp. 230–3, 257), the difficulty has been that in the elementary invariants the total action G and the electromagnetic action $F_{\mu\nu}F^{\mu\nu}$ occur in the combination $G+F_{\mu\nu}F^{\mu\nu}$. In particular, the generalised volume $\sqrt{(-|{}^*G_{\mu\nu}|)}$, which has since been brought into prominence by its use in the Born-Infield theory, reduces to this combination (*loc. cit.*, p. 233). But there seems to be no sense in adding an electromagnetic action to a total action which already includes it. It is the difference $G-F_{\mu\nu}F^{\mu\nu}$, representing material or non-Maxwellian action, which should be represented by the elementary invariants. Devices for changing the sign, e.g. by alternating the suffixes in the invariant ${}^*G_{\mu\nu}{}^*G^{\nu\mu}$, were proposed, but were not very convincing. But the sign is rectified now that we realise that, owing to the identification of the electromagnetic potential with real instead of imaginary changes of gauge, $i\kappa_\mu$ has been substituted for κ_μ and $iF_{\mu\nu}$ for $F_{\mu\nu}$, throughout the field theory as originally given. Consequently $F_{\mu\nu}F^{\mu\nu}$ should have been $-F_{\mu\nu}F^{\mu\nu}$.

With this amendment the field theory as set forth in Chapter VII of *The Mathematical Theory of Relativity* is acceptable today. The investigations in this book have a close connection with it at many points, and confirm it by elucidating the manner in which it forms the macroscopic counterpart of wave mechanics.

After introducing gauge systems transformable at will, Weyl pointed out that there exists at every point of space-time a *natural gauge* furnished by the radius of spherical curvature; and he later reached the conclusion that our actual measures are made in terms of this gauge. This was extended by the writer who showed that a natural gauge, not only at every point but for measurement in every direction at that point, is provided by the contracted curvature tensor, and that the law of gravitation is the expression of the fact that it is to this gauge that our actual macroscopic measurements refer.

By starting with no determinate gauge system, and thereby discovering the natural gauge instead of merely postulating it, Weyl had made a fundamental advance. But, in a sense, his conclusion stultified his premises. The principles of physical measurement are bound up with the natural gauge; we cannot employ alternative gauges without giving to the words "length", "volume", etc. meanings which they do not bear in physics. Gauge transformation had become one of those etymological transformations—which too frequently mar theoretical discussions—proclaiming the obvious truth that if you alter the meanings of words you may assert anything you like. In particular, one of the most attractive features of his electromagnetic theory had to be given up, viz. that the arbitrary gradient, which can be

† *Mathematical Theory of Relativity*, Chapter VII.

added to the electromagnetic potential without altering anything observable, represented the arbitrariness of gauge. Thus Weyl's two results (1) the discovery of variable gauge, which accounted for the existence of quantities which might be identified with electromagnetic potentials, and (2) his discovery of natural gauge, which leads ultimately to the explanation of the law of gravitation, seemed to be contradictory; and it was necessary to suppose that (2) superseded (1). But we can now accept them both, with the modification that the variability referred to in (1) is an "imaginary gauge transformation"; that is to say, it is not a change of the real part of $\log\lambda$ which furnishes the standard for the measurement of lengths and distances, but of the imaginary part of $\log\lambda$, which (although called a gauge transformation by analogy) does not affect the reckoning of length.

In our present development natural gauge is used from the beginning, because displacement first arises as an angular quantity (angle of a transformation) which is the ratio of the linear displacement to the radius of curvature. Thus we do not encounter the preliminary ambiguity which leaves the measure of the displacement indeterminate until the radius of curvature is brought in as standard.† There is no provision in our theory for real change of gauge—for using any other standard. The reality conditions for rotations restrict the transformation (8·81) to imaginary gauge transformations, i.e. real changes of the phase angle θ_{16}.

8·9. Non-integrable Transformations.

In a general way we can trace the origin of the non-integrable gauge transformation which creates an electromagnetic field. A non-integrable transformation arises when we contemplate a field of transformation composed of transformations which do not commute. For example, the transformation $\psi' = e^{E_1\alpha_1 dx_1 + E_2\alpha_2 dx_2}\psi$ is non-integrable. If we apply it to a square circuit composed of successive displacements $(dx_1, 0)$, $(0, dx_2)$, $(-dx_1, 0)$, $(0, -dx_2)$, we obtain as far as the second order

$$\psi' = (1 - E_2\alpha_2 dx_2 - \tfrac{1}{2}\alpha_2{}^2 dx_2{}^2)(1 - E_1\alpha_1 dx_1 - \tfrac{1}{2}\alpha_1{}^2 dx_1{}^2) \times (1 + E_2\alpha_2 dx_2 - \tfrac{1}{2}\alpha_2{}^2 dx_2{}^2)(1 + E_1\alpha_1 dx_1 - \tfrac{1}{2}\alpha_1{}^2 dx_1{}^2)\psi,$$

which reduces to $$\psi' = (1 - 2E_1E_2\alpha_1\alpha_2 dx_1 dx_2)\psi.$$

Thus the result of taking ψ round the circuit is to transform it to

$$\psi' = e^{-2E_{12}\alpha_1\alpha_2 dx_1 dx_2}\psi. \qquad (8\cdot91)$$

This, however, does not immediately solve the problem of the creation of an electromagnetic field, which depends on a similarly non-integrable *algebraic* transformation.

† We are able to start in this way because we treat a very simple uniform space-time, whereas the field theory is concerned with the origin of the natural gauge system in irregular macroscopic space-time.

The field κ_μ is due to systems extraneous to the particle or system S to which the wave vector ψ belongs. As explained at the end of § 2·9 the external particles have their own symbolic frames F_μ, etc., which commute with the E-symbols of the system S. The effect of recognising these external particles will be to introduce into our calculus a large number of additional symbols which commute with the E_μ but not in general with one another; so that a much wider variety of transformations q can be contemplated.

We have hitherto ignored the external particles and the transformations representing relative displacement of them, because the elementary equations suppose S to be in a standard environment, namely neutral space-time. But an electromagnetic field presupposes a non-uniform environment. A change of position dx_μ is not merely a transformation from one point to an equivalent point of space-time (§ 4·4); it involves also an intrinsically different environment of S. Thus the displacement dx_μ will involve a supplementary transformation, representing the change of environment, which we may take to be of the form

$$\psi' = \exp\{Y_1\alpha_1 dx_1 + Y_2\alpha_2 dx_2 + Y_3\alpha_3 dx_3 + Y_4\alpha_4 dx_4\}.\psi, \qquad (8\cdot92)$$

where the Y_μ are composed of symbols belonging to the extraneous systems, and therefore commuting with the E_μ. We do not suppose that the Y_μ are anticommuting symbols of a complete set; they will usually be complicated symbolic expressions. But provided that they imperfectly commute (as normally happens with complicated symbolic expressions), the transformation (8·92) will be non-integrable; and ψ, after being taken round a circuit, will not return to its original value but will undergo a Y rotation of some kind. But since the Y-symbol of the rotation commutes with all the E_μ, it will be indistinguishable from an algebraic transformation; and it will count as an algebraic transformation so far as the E-frame is concerned.

Thus the effect of irregularity of distribution of the surrounding protons and electrons, which might be particularised with almost an infinitude of detail by introducing their own symbolic frames, is reduced to a non-integrable algebraic transformation of the vectors of the E-frame. This transformation represents the difference between the standard environment of neutral space-time and the modified environment—a difference which is recognised as the macroscopic electromagnetic field due to the specialised distribution of the external charges. As we have seen in § 8·8, a field of non-integrable algebraic transformation is equivalent to the insertion of electromagnetic potentials κ_μ in the momentum vector.

We have supposed that the supplementary transformation (8·92) contains only the symbols belonging to the external systems. Would it not be more natural to suppose that it contains combinations of these external symbols with the symbols of S, e.g. $E_\mu Y_\nu$? In that case the non-integrable trans-

formation will not be algebraic in the E-frame, and the field cannot be represented by a potential vector with algebraic components κ_μ. I agree that it would be more natural. But the question for us is, not what actually happens, but *what is supposed to happen* in the ideal problems to which Dirac's equation (8·262) is applied. Actually a charged particle polarises the surrounding distribution of electric charges. When it is displaced the potential due to surrounding charges is altered by their changed polarisation. This effect (Debye-Hückel effect) is of great practical importance. But it is not included in Dirac's equation, which postulates that the field is due to a rigid distribution of charge. If Dirac's equation is applied to an electron in a field in which the Debye-Hückel effect is large, it gives an incorrect value of the energy.

Thus in treating the origin of the κ_μ, i.e. of the electromagnetic terms in Dirac's equation or equivalently the electromagnetic terms in the momentum and energy, we must adhere to the same idealised conditions. The postulate is that the system S itself has no share, direct or indirect, in determining κ_μ. We must therefore omit the terms, if any, which are not invariant for rotations of the particle, substitution of particles of opposite sign† or opposite spin, etc. The terms admitted therefore correspond to a purely algebraic transformation. There is no need for us to show that the omitted terms are small in practical problems; very often they are not.

The term κ_μ in the wave equation (8·262) is essentially a *macroscopic* electromagnetic potential. In microscopic problems the field due to one or more individual particles requires a more complex specification by means of multiple matrices. The criterion is that, if the distribution of particles producing the field can be treated as rigid, (8·262) suffices. The foregoing discussion makes it clear that the field due to the particle itself is not to be included in κ_μ. Neglect of this condition has led to the occurrence of an infinite self-energy of the particle in certain theories.

The internal wave equation for the hydrogen atom, adopted in (9·221), provides an exception to the rule that κ_μ is a macroscopic potential. The equation is of the form (8·262) notwithstanding that the electromagnetic field is due to a single particle (the proton). This is because the problem is transformed by the use of relative coordinates into the motion of a particle in a *rigid* field. It must be emphasised that this is a quite exceptional use of κ_μ, made possible by the simplicity of the problem, and that the microscopic interactions of particles cannot usually be represented by a field of this form.

† There is an apparent change of sign of the electromagnetic terms in the wave equation when a proton is substituted for an electron; but what has really happened is that the electromagnetic terms are unaltered, and all the other terms have changed sign.

CHAPTER IX

THE HYDROGEN ATOM

9·1. Steady States.

Before tackling a practical problem, it is appropriate to recapitulate and systematise certain ideas which have appeared in a scattered way in previous chapters.

In the practical application of wave mechanics the central problem is the search for systems which shall be dynamically steady. The phrase "dynamically steady" requires amplification.

There is an almost inevitable ambiguity in the use of the words "electron" and "proton" in the new physics. We say that (*a*) an electron is no longer a particle but a wave, and (*b*) that the waves specify the probability distribution of an electron. Thus the term is applied both to the distribution and to that which is distributed. For definiteness let us call that which occupies any point of the distribution an *electron-point*. We consider then electron-points distributed over a domain of geometrical coordinates x_μ. A displacement dx_μ is a displacement of the electron-point that is contemplated; no dynamical conception is attached to the displacement; it is a transfer of our attention from one electron-point to another. But we can contemplate also a bodily displacement of the whole distribution; such bodily displacement is described as a change ds_μ of a dynamical coordinate s_μ. Here again the displacement may be regarded primarily as a transfer of attention—from one electron-distribution to another, instead of from one electron-point to another. But when, by habit, we introduce dynamical conceptions they are attached to the displacements ds_μ, not to dx_μ. The dynamical electron—the moving entity—is the probability distribution, and its mode of displacement is wave propagation.

We need not confine attention exclusively to bodily translation or rotation of the distribution. We can consider more general sequences of distributions. The general method of specifying a sequence of distributions is by a transformation $q = e^{iWs}$; then a displacement ds signifies the change of distribution which is produced by applying the transformation e^{iWds} to the vectors describing the distribution. If W is an E-number, this is a relativity rotation of the space vectors defining the distribution, and is therefore a displacement without intrinsic change. The corresponding coordinate s will be called a *simple* dynamical coordinate. By allowing W to include differential operators, we obtain a more general type of displacement including deformation, and define a correspondingly *generalised* dynamical coordinate. Each simple dynamical coordinate is closely related to (and frequently confused with)

a geometrical coordinate x_μ or θ_μ, viz. that defining the direction of the bodily displacement or rotation of the distribution; but the generalised dynamical coordinates have no geometrical counterparts. In practice, however, they often have approximate counterparts; for a generalised coordinate usually appears as a slight modification or adaptation of a simple coordinate. For example, a free electron possesses (simultaneously) four simple dynamical coordinates representing bodily displacement of its distribution in four antiperpendicular directions x_1, x_{23}, x_{45}, x_{16}; when the electron is in the electromagnetic field of a nucleus, we have to find four generalised dynamical coordinates to replace these.

Let us consider a system with four dynamical coordinates s_μ. In what circumstances should we describe the fourfold sequence of distributions as *steady*? It would certainly be considered steady if the distributions were all intrinsically similar. But that is unnecessarily stringent, since we cannot make exhaustive observations of every detail of the distribution. The minimum condition is that some recognisable characteristic of the distribution shall be steady, i.e. constant over the domain of dynamical coordinates s_μ. Since the observable characteristics (physical vectors) are space vectors, we require that a complete space vector J determined by the distribution shall be constant over the domain s_μ. That is to say, J must be invariant for the transformations $q = e^{iW_\mu s_\mu}$.

We take J to be factorisable. It would be possible to obtain a steady state by compounding two pure states neither of which is steady. But the combination is not of practical importance unless there is security that the two states remain superposed with the same relative probability factors when external perturbations are admitted. The argument runs: the steady states which we wish to discover are those which behave as units under external perturbations. Unitary character, i.e. purity, is expressed symbolically by a spectral operator. Therefore to bring our symbolism into line with the physical conditions to which it is applied, we must represent the unit states by spectral operators. The latter are idempotent symbols. We consider in particular the idempotent space vector as the simplest element in the symbolism that is thrust upon us. We regard the simple elements of the symbolism and their physical counterparts, not as hypothetically "existing", but as idealisations which owe their importance to the fact that anything more complicated can be, and commonly will be, analysed into these simple elements.

By § 8·5 a sufficient condition that J shall be constant over the domain s, s_1, s_2, s_3 is that its factors ψ, χ^* shall be common eigensymbols of W, U_1, U_2, U_3. It is easily seen that this condition is also necessary.

In practice we assume that W, U_1, U_2, U_3 commute. It is not true that operators which have a common eigensymbol necessarily commute (§ 3·7 (*f*))

but if W, U_1, U_2, U_3 do not commute they will not be constant over the domain of s, s_1, s_2, s_3, their derivatives being given by (8·52). Transformations in which the operational forms are functions of the dynamical coordinates, e.g. $W(s_1, s_2, s_3)$, are not considered in wave mechanics at present. They bear the same kind of relation to constant transformations that general relativity transformations bear to those of special relativity theory. The domain of such transformations will have a curvature embodying the non-commutability of the rotations in it. Whether it would be profitable to pursue the study of non-commuting symbols with a common eigensymbol, I cannot say. But it may be worth noticing that existing methods of search for steady systems, i.e. distributions with a recognisable characteristic which is constant over a multi-dimensional domain, are not necessarily exhaustive; and it is just possible that steady states, which may have some physical importance, have escaped our analysis.

Limiting ourselves accordingly to constant transformations, *the problem of finding dynamically steady states resolves itself into the finding of four commuting symbols*, or whatever number of symbols may be appropriate to the kind of system investigated. The symbols represent constant characteristics of the system; but they are generally portions of the constant space vector J, and do not imply any constancy of the system additional to that originally postulated. For example, when the operators are E-symbols, $W+U_1+U_2+U_3=E_{16}+E_1+E_{23}+E_{45}=J$.

We must next try to understand why these "steady states" are important. Only one of the four coordinates is conceived as time displacement. Our dynamical picture of a system pursuing a trajectory in the domain of s, s_1, s_2, s_3 does not suggest any reason why J should be required to be constant in directions transverse to the trajectory. The importance of the latter condition is that it introduces the maximum degeneracy into statistical enumerations. In accepting J as the criterion of steadiness, we implicitly decide to count all configurations which have the same J as one configuration. Thus the probability occupying the whole of S-space counts as the probability of one configuration. Our picture of a configuration as an isolated point in S-space pursuing a trajectory does not apply; it is a whole continuum or wave front that "travels". Moreover, when (as in quantised systems) the four-dimensional S-space exists only for discrete values of J, and intermediate values of J occupy loci of three or fewer dimensions, the discrete values have infinitely greater probability than the intermediate values.

An S-space formed by generalised dynamical coordinates is not on quite the same footing as an S-space formed by simple dynamical coordinates. The difference is that the generalised displacement is a transformation peculiar to the system, and is not applicable to its idealised environment.

Now a system without an environment is unthinkable; and it is no use displacing it all over the S-space if it cannot take with it the environment which its structure demands. That was the early mistake of relativity theory, which applied transformations to the differential equations but omitted to provide for their application to the boundary conditions. The standard environment is uniform, i.e. spherical, neutral space-time. This is conceived as permanent; so that the only transformations admitted are those which transform it into itself, namely the kinematical E-rotations. Hence, in general, if we apply generalised W transformations to a system, it will no longer fit the boundary conditions where it merges into standard space-time. Physically we should say that the new configuration requires a pressure or an electromagnetic field to maintain it.

It might seem from this that the importance of generalised steady states is fictitious. But we have to remember that the standard environment is a simplification of the actual environment. The practical physicist is not concerned with a hydrogen atom existing alone in uniform spherical space. He deals with hydrogen atoms surrounded by other atoms or ions, or in fields of radiation. We admit that it would be useless to consider generalised displacements of a system which could not be applied to its environment; but this is discounted by the fact that we cannot say beforehand what the exact environment will be, and what type of displacement it will admit. We therefore investigate the most general steady states, since they may become realisable if the environment is appropriate.

To conform to our observational knowledge there should be just one steady state of a hydrogen atom in the standard environment, namely the ground state. The other states can be realised if the environment is such as is capable of "exciting" the atom—in particular if it contains a field of radiation.

We proceed to an investigation of the steady states of a hydrogen atom. This belongs to quantum theory proper, and strictly speaking is outside our territory. But it is necessary for *liaison* purposes to follow through in our own notation, and from our own point of view, one practical problem which introduces the leading ideas of quantum theory. Unexpectedly the investigation has proved to be vitally important, because it reveals an inconsistency of a factor 2 (for which, I think, the quantum physicists must be held responsible) which led to an error in the numerical results found in earlier versions of the present theory.

9·2. The Commuting Operators for a Hydrogen Atom.

The conditions for a steady internal state of a hydrogen atom were obtained approximately by Schrödinger as a differential equation of the second order. This is now superseded by Dirac's exact treatment. It is commonly said that

Dirac replaced Schrödinger's second order equation by a first order equation $(W-m)\psi=0$. I regard this as a misconception. The modern equivalent of Schrödinger's equation is the set of partial differential equations $\partial W/\partial s_\mu=0$, $\partial U_1/\partial s_\mu=0$, etc., which secure that certain characters remain steady for fourfold displacement. Since W, U_1, U_2, U_3 contain differential operators, these equations are of the second order. The introduction of E-matrices has not affected the fundamental conclusion that *the condition that ψ shall represent a steady state is expressed by differential equations of the second order.*

By the dynamical equations (8·52) the equation $\partial W/\partial s_1=0$ is equivalent to $U_1W-WU_1=0$. In this way the original set of equations is reduced to the condition

$$W,\ U_1,\ U_2,\ U_3 \text{ mutually commute.} \qquad (9\cdot21)$$

We may regard (9·21) as a first integral of the second order equations.

The way in which Dirac's wave equation enters into the problem is that it fixes the analytical form of one of the symbols W in (9·21); and hence it is the starting point for determining three other analytical forms which commute with it and with one another.

We have shown in § 9·1 that ψ must be a common eigensymbol of W, U_1, U_2, U_3; so that, if m is the eigenvalue of W, we have $(W-m)\psi=0$. This equation is provided by (8·261), which we found as a condition for the conservation of probability. We could not at that time show that m was constant in time; but this now follows from the dynamical equations, because W is constant. We have still to show that m is the same for different states of the system. It will be proved in Chapter XII that (for particles represented by simple wave tensors) m has one of two absolutely determined values, corresponding respectively to electrons and protons. Using this result in anticipation, we shall take m to be a constant of nature.

The eigenvalue m of W is imposed from outside; there are no limitations on the eigenvalues of U_1, U_2, U_3 other than those which will be found in the course of the investigation. This difference is due to the fact that s is *singled* out as the connecting link with extraneous systems (§ 8·5).

We now consider the wave equation (8·262) for an electron in an electrostatic field of positive potential proportional to $1/r$, so that

$$\kappa_\mu=(0,\ 0,\ 0,\ -i\alpha/r), \quad r^2=x_1^2+x_2^2+x_3^2.$$

Then writing (8·262) in the form $(W-m)\psi=0$, we have

$$W=-i\left(E_1\frac{\partial}{\partial x_1}+E_2\frac{\partial}{\partial x_2}+E_3\frac{\partial}{\partial x_3}+E_4\frac{\partial}{\partial x_4}+E_4\frac{\alpha}{r}\right). \qquad (9\cdot221)$$

We seek three other operators, linear in $\partial/\partial x_\mu$, which commute with W and

with one another. It is easily verified that the following satisfy this condition:

$$U_1 = iE_4\left\{E_{23}\left(x_3\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_3}\right) + E_{31}\left(x_1\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_1}\right) + E_{12}\left(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}\right) - 1\right\}, \quad (9\cdot222)$$

$$U_2 = i\left(x_3\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_3} + \tfrac{1}{2}E_{23}\right), \quad (9\cdot223)$$

$$U_3 = \frac{\partial}{\partial x_4} = -i\frac{\partial}{\partial t}. \quad (9\cdot224)$$

We denote the eigenvalues of U_1, U_2, U_3, W by μ, u, ϵ, m, and our problem is to find the relations which must exist between them, in order that the common eigensymbol ψ may satisfy appropriate boundary conditions. More particularly we wish to determine ϵ, which is the energy conjugate to t, in terms of m, μ, u. The method of solution here followed is mainly due to Temple.†

We have enunciated the problem in an abstract way, without mentioning the hydrogen atom. But (9·221) is the hamiltonian of the hydrogen atom adopted in Dirac's theory, and we shall provisionally assume that that is the physical application. There is no obvious reason to expect that the wave equation (8·262), in which the κ_μ are potentials of a macroscopic electric field (§ 8·9), will be adequate in the interior of an atom; so that for the time being the term $E_4\alpha/r$ in (9·221) is only justified empirically. But in Chapter XV we shall find that (9·221) is the *exact* hamiltonian for a system of two elementary particles and we shall determine the theoretical value of the constant α.

9·3. Solution of the Equations.

Let
$$E_r = (E_1x_1 + E_2x_2 + E_3x_3)/r \quad (9\cdot311)$$
so that $E_r^2 = -1$, and E_r anticommutes with E_4. By direct multiplication

$$rE_r\left(E_1\frac{\partial}{\partial x_1} + E_2\frac{\partial}{\partial x_2} + E_3\frac{\partial}{\partial x_3}\right) = -\left(x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2} + x_3\frac{\partial}{\partial x_3}\right) + E_{23}\left(x_2\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_2}\right) + E_{31}\left(x_3\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_3}\right) + E_{12}\left(x_1\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_1}\right)$$
$$= -r\partial/\partial r - iE_4U_1 - 1. \quad (9\cdot312)$$

Hence by (9·221)
$$E_rW = i\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) - \frac{E_4U_1}{r} - iE_rE_4\left(U_3 + \frac{\alpha}{r}\right). \quad (9\cdot313)$$

Multiplying this by the common eigensymbol ψ, the operators W, U_1, U_3 reduce to their eigenvalues m, μ, ϵ, so that

$$\left\{i\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) - \frac{E_4\mu}{r} - iE_rE_4\left(\epsilon + \frac{\alpha}{r}\right) - E_rm\right\}\psi = 0. \quad (9\cdot321)$$

† *Proc. Roy. Soc.* A, 127, 349 (1930).

Or, writing $\psi = r^{-1}\omega$,

$$\left\{i\frac{\partial}{\partial r} - E_4\frac{\mu}{r} - iE_rE_4\left(\epsilon + \frac{\alpha}{r}\right) - E_r m\right\}\omega = 0. \tag{9·322}$$

Let $$F = -E_rE_4\epsilon + iE_r m, \quad G = iE_4\mu - E_rE_4\alpha, \tag{9·331}$$

so that the equation is $$\frac{\partial\omega}{\partial r} + \left(F + \frac{G}{r}\right)\omega = 0. \tag{9·332}$$

By (9·331) $$F^2 = f^2, \quad G^2 = g^2, \quad FG + GF = -2\alpha\epsilon, \tag{9·333}$$

where $$f = (m^2 - \epsilon^2)^{\frac{1}{2}}, \quad g = (\mu^2 - \alpha^2)^{\frac{1}{2}}. \tag{9·334}$$

To solve (9·332) we make the algebraic substitution

$$r = x/2f, \quad \omega = e^{-\frac{1}{2}x}\xi, \tag{9·341}$$

and take ξ and x as our new variables. The equation becomes

$$\frac{\partial\xi}{\partial x} + \frac{G}{x}\xi + \tfrac{1}{2}\left(\frac{F}{f} - 1\right)\xi = 0. \tag{9·342}$$

In an ordinary algebraic equation this transformation would have removed the last term; but in the symbolic equation it leaves a *singular* coefficient $\frac{1}{2}(F/f - 1)$ instead of a zero coefficient. Assume a solution in series

$$\xi = \sum_0^\infty C_n x^{n+p}, \tag{9·351}$$

where the C_n may be non-algebraic. In order that the integral of the three-dimensional density $\int\psi\phi^* r^2 dr$ may be finite for a region enclosing the origin, ψr and $\phi^* r$ must be finite or else tend to infinity less rapidly than $r^{-\frac{1}{2}}$. Hence ω is finite or diverges less rapidly than $r^{-\frac{1}{2}}$, so that

$$p > -\tfrac{1}{2}. \tag{9·352}$$

Further, in order that the same integral may converge when the region extends to infinity, the series must terminate at some finite value of n, say $n = n'$. (It can be proved that, unless the series terminates, it diverges when $x \to \infty$.)

Substituting (9·351) in (9·342) and equating coefficients of x^{n+p-1}, we obtain the recurrence relation

$$(n + p + G)C_n = -\tfrac{1}{2}(F/f - 1)C_{n-1}. \tag{9·361}$$

Setting $n = 0$, $(p + G)C_0 = 0$. So that C_0 is an eigensymbol of G, and $-p$ is an eigenvalue of G. Hence, by (9·333), $p = \pm g$. We shall find later that there is no possible value of g between 0 and $\frac{1}{2}$, so that by (9·352) $p = g$.

Setting $n = n' + 1$, we obtain $\frac{1}{2}(F/f - 1)C_{n'} = 0$; so that $C_{n'}$ is an eigensymbol of F, the eigenvalue being f.

Multiply (9·361) by $(F/f + 1)$; since $F^2 = f^2$, we have

$$(F/f + 1)(n + g + G)C_n = 0. \tag{9·362}$$

Hence, by (9·333),

$$\{(n + g)(F/f + 1) + G(1 - F/f) - 2\alpha\epsilon/f\}C_n = 0. \tag{9·363}$$

Now set $n=n'$. Since $FC_{n'}=fC_{n'}$, we obtain

$$n'+g-\alpha\epsilon/f=0. \tag{9·371}$$

Or, by (9·334),

$$\frac{\epsilon}{\sqrt{(m^2-\epsilon^2)}}=\frac{n'+\sqrt{(\mu^2-\alpha^2)}}{\alpha}. \tag{9·372}$$

This is Sommerfeld's formula for the energy ϵ. Here n' is a positive integer or zero. It remains to determine the possible values of μ.

9·4. The Eigenvalues of U_1, U_2.

Take spherical polar coordinates r, θ, ϕ, so that ϕ is the azimuthal angle in the plane x_2x_3. Then $U_2=-i\partial/\partial\phi+\frac{1}{2}iE_{23}$; and, since u is its eigenvalue,

$$(U_2-u)\psi=(-i\partial/\partial\phi+\tfrac{1}{2}iE_{23}-u)\psi=0. \tag{9·411}$$

Let

$$\chi=e^{-(iu+\frac{1}{2}E_{23})\phi}\psi. \tag{9·412}$$

Then

$$\frac{\partial\chi}{\partial\phi}=e^{-(iu+\frac{1}{2}E_{23})\phi}\left(-iu-\tfrac{1}{2}E_{23}+\frac{\partial}{\partial\phi}\right)\psi$$

$$=0 \quad \text{by (9·411).}$$

Accordingly χ is constant for change of ϕ; and since ψ (*assumed to be single-valued*) is unaltered when ϕ is increased by 2π, we have

$$\psi=e^{-(iu+\frac{1}{2}E_{23})2\pi}\psi.$$

Or, since

$$e^{E_{23}\pi}=-1=e^{i\pi},$$

$$\psi=e^{-i(u+\frac{1}{2})2\pi}\psi. \tag{9·413}$$

Therefore $u+\frac{1}{2}$ is a positive or negative integer. The eigenvalues of U_2 are therefore the half odd integers positive and negative.

To find the possible values of μ, we write

$$\omega_1=x_3\partial/\partial x_2-x_2\partial/\partial x_3+\tfrac{1}{2}E_{23}, \text{ etc.} \tag{9·421}$$

Then it is easily verified that

$$\omega_1\omega_2-\omega_2\omega_1=\omega_3, \text{ etc.} \tag{9·422}$$

Hence

$$\omega_1(\omega_2+i\omega_3)=(\omega_2+i\omega_3)(\omega_1-i), \tag{9·431}$$

$$\omega_1(\omega_2-i\omega_3)=(\omega_2-i\omega_3)(\omega_1+i), \tag{9·432}$$

$$(\omega_2+i\omega_3)(\omega_2-i\omega_3)=\omega_2^2+\omega_3^2-i\omega_1, \tag{9·433}$$

$$(\omega_2-i\omega_3)(\omega_2+i\omega_3)=\omega_2^2+\omega_3^2+i\omega_1. \tag{9·434}$$

By (9·222) and (9·223)

$$iE_4U_1=E_{23}\omega_1+E_{31}\omega_2+E_{12}\omega_3+\tfrac{1}{2}, \quad U_2=i\omega_1. \tag{9·441}$$

Since E_4 commutes with U_1, we obtain by squaring the first expression,

$$U_1^2=-\omega_1^2-\omega_2^2-\omega_3^2+2iE_4U_1-\tfrac{3}{4},$$

using (9·422). Hence

$$\omega_2^2+\omega_3^2=-(U_1-iE_4)^2+\tfrac{1}{4}+U_2^2 \tag{9·442}$$

and, by (9·434),

$$(\omega_2 - i\omega_3)(\omega_2 + i\omega_3) = -(U_1 - iE_4)^2 + (\tfrac{1}{2} + U_2)^2. \qquad (9{\cdot}443)$$

Our solution ψ is an eigensymbol of U_1, U_2, but not of E_4. We can, however, obtain a common eigensymbol χ_0 of the three commuting symbols U_1, U_2, E_4, with eigenvalues μ, u, $-i$, by the method of § 3·7 (*e*), namely $\chi_0 = (E_4 - i)\psi$, since this gives $(E_4 + i)\chi_0 = 0$. Hence by (9·442) and (9·443)

$$(\omega_2^2 + \omega_3^2)\chi_0 = \{-(\mu - 1)^2 + \tfrac{1}{4} + u^2\}\chi_0, \qquad (9{\cdot}451)$$

$$(\omega_2 - i\omega_3)(\omega_2 + i\omega_3)\chi_0 = \{-(\mu - 1)^2 + (\tfrac{1}{2} + u)^2\}\chi_0. \qquad (9{\cdot}452)$$

We introduce a series of symbols defined by the recurrence relation

$$\chi_r = (\omega_2 + i\omega_3)\chi_{r-1}, \quad \chi_{-r} = (\omega_2 - i\omega_3)\chi_{-r+1}, \qquad (9{\cdot}46)$$

r being positive. Multiplying (9·452) initially by $(\omega_2 + i\omega_3)$, it becomes

$$(\omega_2 + i\omega_3)(\omega_2 - i\omega_3)\chi_1 = \{-(\mu - 1)^2 + (\tfrac{1}{2} + u)^2\}\chi_1. \qquad (9{\cdot}471)$$

Also

$$i\omega_1\chi_1 = i\omega_1(\omega_2 + i\omega_3)\chi_0 = i(\omega_2 + i\omega_3)(\omega_1 - i)\chi_0$$

by (9·431). Since $U_2 = i\omega_1$, this gives

$$U_2\chi_1 = (\omega_2 + i\omega_3)(u + 1)\chi_0 = (u + 1)\chi_1. \qquad (9{\cdot}472)$$

Adding (9·471) and (9·472) and using (9·433)

$$\begin{aligned}(\omega_2^2 + \omega_3^2)\chi_1 &= \{-(\mu - 1)^2 + (\tfrac{1}{2} + u)^2 + (u + 1)\}\chi_1 \\ &= \{-(\mu - 1)^2 + \tfrac{1}{4} + (u + 1)^2\}\chi_1. \qquad (9{\cdot}473)\end{aligned}$$

By (9·472) the substitution of χ_1 for χ_0 changes the eigenvalue of U_2 from u to $u_1 = u + 1$; and, comparing (9·473) with (9·451), we see that the eigenvalue of $(\omega_2^2 + \omega_3^2)$ retains the same form with u changed to u_1.

Proceeding step by step, we find that for χ_r, the eigenvalues of U_2 and $\omega_2^2 + \omega_3^2$ are

$$u_r = u + r, \quad \{-(\mu - 1)^2 + \tfrac{1}{4} + u_r^2\}. \qquad (9{\cdot}474)$$

Since $U_2 = i\omega_1$, the eigenvalues of ω_1^2 are negative. Similarly the eigenvalues of ω_2^2 and ω_3^2 are all negative. If this implies that the eigenvalues of $\omega_2^2 + \omega_3^2$ are all negative,† there is (for given μ) an upper limit to the value of u_r^2 in (9·474). Hence the series of symbols χ_r must terminate with symbols χ_k, $\chi_{-k'}$, so that

$$\chi_{k+1} = (\omega_2 + i\omega_3)\chi_k = 0, \quad \chi_{-k'-1} = (\omega_2 - i\omega_3)\chi_{-k'} = 0. \qquad (9{\cdot}481)$$

Adapting (9·452) to χ_k, the condition gives

$$0 = -(\mu - 1)^2 + (\tfrac{1}{2} + u_k)^2.$$

We find similarly $0 = -(\mu - 1)^2 + (-\tfrac{1}{2} + u_{-k'})^2$, so that

$$(\mu - 1)^2 = (u + k + \tfrac{1}{2})^2 = (u - k' - \tfrac{1}{2})^2. \qquad (9{\cdot}482)$$

Hence, for a fixed value of μ, the possible values of u range between $|\mu - 1| - \tfrac{1}{2}$ and $-|\mu - 1| + \tfrac{1}{2}$, the former limit corresponding to $k = 0$ and the latter to $k' = 0$.

† The legitimacy of this inference is examined later.

For our purpose the pertinent result is that by (9·482), μ is a positive or negative integer; for we have already seen that $u+\frac{1}{2}$ is a positive or negative integer. We are therefore able to calculate the energies ϵ of all possible steady states by giving n' and μ integral values in (9·732). Zero value of μ is excluded because it would make the term $\sqrt{(\mu^2-\alpha^2)}$ imaginary; zero value of n' is not excluded.

The assumption made above that $\omega_2{}^2+\omega_3{}^2$ has only negative eigenvalues requires consideration. I do not think there is a general law for non-commuting symbols that if the eigenvalues of X and Y are all negative the eigenvalues of $X+Y$ are all negative. But here we contemplate a restricted type of eigensymbol ψ. In one form or another conditions must be introduced which make the expectation value with respect to ψ intermediate between the greatest and least eigenvalues. As this part of the theory of eigenfunctions in wave mechanics does not concern us very closely, we shall not enter upon it. Accepting such conditions, the expectation values of $\omega_1{}^2$ and $\omega_2{}^2$ will always be negative, and therefore the expectation value of $\omega_1{}^2+\omega_2{}^2$ will always be negative. The eigenvalues, being particular cases of expectation values, will therefore be negative.

9·5. Metastable States.

According to the theory which will be developed in Chapter XV, our adopted wave equation for the hydrogen atom is *exact*; that is to say, the interaction of the proton and electron is precisely expressed by a potential $\kappa_4=-i\alpha/r$. There is no failure of the formula however small r may be. We have found exact solutions of the exact equations; and they agree with the well-known series of quantum states of the hydrogen atom. But we seem to have proved too much! Observationally these states are only imperfectly steady; even if the atom is undisturbed, they do not endure indefinitely. Thus a state which satisfies the exact theoretical conditions for a steady state is found observationally to be imperfectly steady.

The explanation is that in practical applications we have to take into consideration the environment as well as the atom itself (p. 143). We have found the exact solutions of the differential equations; but whether a particular solution has an exact counterpart in nature depends on whether the boundary conditions which it demands are forthcoming. The boundary conditions of a "state" are difficult to visualise. Our general outlook is that whatever exists outside† the state is to be treated as a possible source of perturbations; an environment is therefore regarded as conformable to the boundary conditions of the state if it causes no perturbation of the state.

Light is thrown on this subject by distinguishing between the algebraic and the non-algebraic wave functions of a hydrogen atom. The algebraic

† In the sense of not belonging to the state, not necessarily exterior to it in space.

solutions of (9·332) are easily found. If ω is an algebraic wave function, the term $\partial/\partial r$ in

$$(\partial/\partial r + F + G/r)\,\omega = 0$$

reduces to an algebraic function of r, so that ω must be an eigensymbol of $F + G/r$ for all values of r. Therefore ω is an eigensymbol of F and G. Then by § 3·7 (f), $FG - GF$ is singular (or zero). By (9·331)

$$\tfrac{1}{2}(FG - GF) = iE_r\epsilon\mu - E_rE_4m\mu + iE_4m\alpha, \tag{9·51}$$

so that

$$\tfrac{1}{4}(FG - GF)^2 = \epsilon^2\mu^2 - m^2\mu^2 + m^2\alpha^2, \tag{9·52}$$

since E_r, E_4, E_rE_4 anticommute. The right-hand side must vanish, because $FG - GF$ can have no reciprocal. We have therefore

$$\epsilon^2/(m^2 - \epsilon^2) = (\mu^2 - \alpha^2)/\alpha^2. \tag{9·53}$$

Comparing with (9·372), we see that the algebraic wave functions correspond to $n' = 0$. In this case the series (9·351) reduces to its first term, and the wave function is

$$\omega = C_0 e^{-\frac{1}{2}x}x^{\sigma}. \tag{9·54}$$

It is well known that the states given by $n' = 0$ are the metastable states. Accordingly the distinction between the metastable and the unstable states is that the former have algebraic wave functions and the latter non-algebraic wave functions.

It appears therefore that, in order to satisfy the boundary conditions furnished by the standard environment —uniform neutral space-time—the wave function must be algebraic. Non-algebraic wave functions do not precisely satisfy the boundary conditions; but the discrepancy can be treated as a perturbation. The spontaneous transitions from these states to lower states are attributable to the perturbations which represent this discrepancy. Even the algebraic wave functions are not quite perfectly conformable to the boundary conditions, the only really permanent state being the ground state.

9·6. Single-valuedness of the Internal Wave Function.

We must now refer to the assumption italicised in the second paragraph of § 9·4. In determining the possible eigenvalues of U_2 and U_1, ψ is assumed to be a single-valued function of rectangular coordinates. But the assumption is untrue. Dirac's Lorentz invariant wave vector ψ is necessarily a double-valued function of rectangular coordinates (§ 4·7). Little attention seems to have been paid to this inconsistency; but it reveals a flaw in the foundations of the current theory of the hydrogen atom.

We have to consider two alternatives:

(1) The investigation in § 9·4 must be amended so as to apply to double-valued instead of single-valued wave functions.

(2) The eigensymbol ψ is a single-valued function, and the investigation in § 9·4 is correct; but in this case ψ cannot be a wave vector, and $(W - m)\psi = 0$

cannot be Dirac's original Lorentz-invariant equation. Its origin therefore remains to be investigated.

Our theory will be found to lead to the second alternative; but we may briefly consider the first. The result of taking ψ to be double-valued is that, when the angle ϕ is increased by 2π, ψ may become either $+\psi$ or $-\psi$. We can then deduce that the eigenvalues of u are the integers and half-integers. No further change is made until we reach (9·482), which shows that the eigenvalues of μ will also be the integers and half integers. We obtain therefore twice as many eigenvalues of μ and of u as in the previous discussion. The additional eigenstates do not correspond to observed states of the atom.

Double-valuedness of the eigenfunction has been discussed by Temple.† It is pointed out that there would be a distinction between the integral and the half-integral eigenvalues, the latter being in a sense ineffective because a displacement has no matrix components corresponding to them. It is doubtful whether this ought to be regarded as an excuse for, or an objection to, a theory which employs the double-valued ψ. But the crux of the matter is the exclusion principle. The observational result is that when we consider a nucleus with a number of electrons, the electrons occupy the states given by integral values of μ and $u+\frac{1}{2}$, and ignore the "ineffective" half-integral values. It would therefore be necessary to abandon the accepted form of the exclusion principle, and substitute a new principle according to which only a quarter of the cells, into which phase space is divided by the eigenfunctions, are allowed to be occupied. This would be a very drastic alteration of existing quantum theory.

Following the second alternative, the only change required is to recognise that the eigenfunction ψ of an internal state of the hydrogen atom (distinguished as ψ_H) returns to its original value when the space axes are rotated through 2π, and therefore rotates twice as fast as the wave vector ψ (distinguished as ψ_D) in Dirac's Lorentz-invariant wave equation. Thus ψ_H transforms like a space vector or strain vector. We have seen that the internal configurations of a system are specified by strain vectors, and we therefore take ψ_H to be a strain vector. It must be of index 2, since a strain vector of index 0 (representing a combination of waves travelling in opposite directions) would not have the unidirectional properties of ψ_H.

Current theory has assumed that the momentum in practical units is given by

$$p_\mu \psi_H = -\frac{ih}{2\pi}\frac{\partial}{\partial x_\mu}\psi_H. \qquad (9{\cdot}61)$$

† *An Introduction to Quantum Theory*, pp. 106, 131. His discussion (which is based on a treatment by Born and Jordan) refers to the Schrödinger scalar wave function. It is pointed out that (even in that case) ψ may not be a single-valued wave function, and that the usual inference as to the integral values of the angular quantum numbers is "decidedly precarious".

It is on this basis that the observational value of h is determined, and we must accept it as a definition of h. The momentum operator for a tensor of index 2 being $-ih/2\pi$, the standard momentum operator for a wave vector of index 1 will be $-ih/\pi$ by (8·65). We have therefore

$$p_\mu \psi_D = -\frac{ih}{\pi}\frac{\partial}{\partial x_\mu}\psi_D. \tag{9·62}$$

The common assumption, that the momentum operator which applies to the wave function of a hydrogen atom applies also to the relativistic wave functions introduced in Dirac's Lorentz-invariant equation, is thus found to be untenable. The error of a factor 2 has escaped notice, because the Dirac wave functions are highly abstract (being referred to an unobservable geometrical frame) and are not directly concerned in comparisons of theory and observation; it is in connection with internal wave functions, which (as we have seen) are not Lorentz invariant (§ 7·1), that quantising conditions arise. But it is important that the factor should be set right in fundamental investigations of the connection between relativity theory and quantum theory. We have therefore to note that the theoretical unit of mass used in our fundamental investigations is such that $h=\pi$, not 2π, and that the momentum operator for wave-vector functions is (in practical units)

$$p_\mu = -\frac{ih}{\pi}\frac{\partial}{\partial x_\mu}. \tag{9·63}$$

It is in accordance with the theory of phase space in Chapter VII that the configurations or states of the hydrogen atom should be discriminated by strain vectors of index 2 rather than by wave vectors. The theory of the energy levels of hydrogen thus falls into line with the general theory of the representation of internal configurations of systems. Looking at the matter from a physical point of view, the double-valuedness of ψ_D reflects the abstract character of the analysis which introduces it. We can only observe relations between two systems or parts of a system, so that the minimum we can contemplate observationally is a double function $\psi_D.\psi_D'$. Rotation of the axes through 2π no longer introduces an ambiguity of sign; each factor becomes multiplied by $e^{\pi i}$, and the product by $e^{2\pi i}$. When we consider relative motion, i.e. refer each particle to the other as origin, ψ_D, ψ_D' become equal; ψ_D gives the distribution of the electron relative to the proton, and ψ_D' the distribution of the proton relative to the electron. In this case $\psi_H=\psi_D.\psi_D$.† If we prefer to represent the state by a wave vector ψ_D, we must remember that $\psi_D=\pm\sqrt{\psi_H}$, and not ignore the ambiguity of sign

† In this way wave tensors of index 2 which are perfect squares acquire a special prominence in practical problems. The wave tensors of index 0 which correspond to them are Hermitic (§ 8·6). This seems to be the origin of the Hermitic conditions, which have been applied too indiscriminately in current quantum theory.

which is an essential characteristic of ψ_D. Taking the square root is, however, a gratuitous complication; and the actual investigation of the hydrogen atom is based on ψ_H.

As a working rule, we note that normally the coefficient of the momentum operator will be $-ih/\pi$ for absolute coordinates of a particle referred to a geometrical origin, and $-ih/2\pi$ for its relative coordinates referred to another particle as origin. But it is impossible to lay down a universal rule, since the momentum operator depends on the index of the wave function adopted (8·65).

Accordingly in practical units the special form for the hydrogen atom of the wave equation (8·262) of the double-valued wave function will be

$$\left\{-\frac{ih}{\pi}\left(E_1\frac{\partial}{\partial x_1}+E_2\frac{\partial}{\partial x_2}+E_3\frac{\partial}{\partial x_3}+E_4\frac{\partial}{\partial x_4}\right)-\frac{ie^2}{cr}E_4-m\right\}\psi_D=0, \quad (9\cdot64)$$

since in practical units $\kappa_\mu=(0, 0, 0, -ie^2/cr)$. When, as in (9·221), the operand is $\psi_H=\psi_D.\psi_D$, the coefficient h/π must be replaced by $h/2\pi$ to give the same condition. It follows that the constant α in (9·221) is

$$1/\alpha=hc/2\pi e^2. \quad (9\cdot65)$$

Attention may be called to a further point on which we are unable to accept the current view. In (9·224), following the usual notation, we have denoted the coordinate conjugate to U_3 by t. We must now point out that t *has nothing to do with time.* Our equations refer to the internal states of the atom, and we have seen (§ 7·1) that the only time-variable in an internal state is the dynamical or proper time s which is conjugate to W. If the atom is moving in an external frame the coordinate time in that frame is obviously not the t referred to in the equations. The dynamical coordinate conjugate to U_3 is of an altogether different nature, and will be identified in Chapter XV with the linearised permutation coordinate of the proton and electron.

CHAPTER X

DOUBLE WAVE VECTORS

10·1. Multiplication of Probabilities.

Probabilities are combined by multiplication. We have seen (§ 7·7) that the configuration of a system and the probability of that configuration are specified by a single symbol—a strain vector. The multiplication of probabilities therefore involves a multiplication of the strain vectors; and the treatment of a combined system is based on the multiplication of the vectors descriptive of its separate parts.

If the combined probability of two events is precisely the product of their separate probabilities, the two events are said to be *independent*. Similarly, we define independent systems to be such that the combined vector specifying probability and configuration is the product of the separate vectors. If in combining the systems the product is modified, the modification constitutes an *interaction*. For the present we consider combination without interaction.

It is, of course, entirely opposed to our habit of thought to regard a system as the product of its parts rather than as the sum of its parts. Therefore we have a long way to travel before we can connect the combination of systems by multiplication with our ordinary outlook.

Admitting that multiplication is the primary operation in the theory, the prevalence of exponentials in the formulae which we have developed indicates the way in which the subsidiary operation of addition arises. We see that in general the additive quantities of physics must occur in exponentials. An elementary example is afforded by action, which is well known to be additive; in wave mechanics it is represented as the phase angle θ of an imaginary exponential. Other cases are not so simple. We speak of the density ρ in a unit volume, due to the probability distribution of a particle, without specifying whether it is a low probability of a high mass or a high probability of a low mass; this agrees with the principle that the probability p and the characteristic m of the system are not to be detached from one another. If we consider a second particle, the combination of probabilities pp' involves taking the product $pmp'm' = \rho\rho'$. But the combined system is ordinarily considered to be characterised by a density $\rho + \rho'$. This is a point to be investigated in due course.†

In so far as an addition $\alpha + \beta$ arises from a multiplication $e^{\alpha}.e^{\beta}$, addition must be looked upon as a possibly non-commutative operation. We have to

† The methods by which wave mechanics is adapted to treat additive properties instead of multiplicative properties are explained in Chapter XVI.

treat "ordered sums" of matrices as we treat ordered products. An integral may also require to be ordered. I do not propose to use the algebra of non-commutative addition, which is even less familiar than the non-commutative multiplication which we have employed; I mention it in order to show the type of mathematical complication in which we should be involved if we attempted to follow the common outlook more closely.

Our first task is to study the double wave tensors which are formed when we multiply the simple wave tensors specifying the probabilities and configurations of two independent systems.

10·2. Double Frames.

By multiplying two wave vectors ψ_α, ϕ_β belonging to two independent systems we obtain a 16-valued quantity $\Psi_{\alpha\beta}$ which we call a double wave vector. Formally it resembles a wave tensor of the second rank (mixed or covariant) obtained by multiplying wave vectors ψ_α, χ_β belonging to the same system; but it has a wider field of transformation. In a wave tensor a transformation of χ is locked to the transformation of ψ; but in a double wave vector there is no such restriction. If T is a covariant wave tensor and Ψ a double wave vector, their most general transformations are

$$T_{\alpha\beta}' = q_{\alpha\gamma} q_{\beta\delta} T_{\gamma\delta}, \quad \Psi_{\alpha\beta}' = q_{\alpha\beta\gamma\delta} \Psi_{\gamma\delta}, \tag{10·21}$$

the former involving 16 coefficients $q_{\alpha\beta}$, and the latter 256 coefficients $q_{\alpha\beta\gamma\delta}$.

For example, let the objects described by the wave vectors ψ, ϕ be the earth and moon. We may use different axes of reference for the momenta, spins, etc., of the earth and moon. If we change the axes, there is no need to apply the same transformation to those used for the earth and those used for the moon. It is true that we do not usually avail ourselves of this liberty, and a transformation of axes would ordinarily apply to both bodies; but that is because we contemplate an earth-moon system, which is not merely a mental association of the earth and moon, but comprises physical relations between the two bodies not to be found in either body separately.

More usually we are interested in the other aspect of a relativity transformation, in which it is regarded as a displacement without intrinsic change of a physical system, the frame being kept fixed. In the case of the earth and moon, this implies the introduction of two sets of vectors ψ_α, ϕ_β capable of being rotated independently.

To provide for independent transformations of ψ and ϕ we must introduce two symbolic frames. We have therefore two complete orthogonal sets E_μ, F_μ. In the matrix representation of the symbols we take (for convenience) E_μ and F_μ to correspond to the *same* matrix; but in chain multiplication E_μ is connected with the first suffix of a preceding or following double vector,

and F_μ with the second. If more than one E or F symbol appears in a product, the E's are in chain with the E's and the F's with the F's. The order of writing E's and F's is of no importance; that is to say, every E_μ commutes with every F_μ. As usual $E_{16}=i$, $F_{16}=i$.

Unless the contrary is stated we shall suppose that E_μ and F_μ are right-handed sets. It may sometimes be more appropriate to use a right-handed and a left-handed set; in this case the matrices E_μ and F_μ cannot be identical, F_{16} and one of the pentads being reversed in sign as compared with the corresponding E-matrices. The corresponding modification of the various formulae is easily found.

Since (for right-handed sets) the letters E, F merely indicate the chain connection, there is no need to distinguish them when row-and-column suffixes are inserted. The EF notation is a device for extending the matrix notation, which omits suffixes, to expressions of the fourth rank such as $q_{\alpha\beta\gamma\delta}$.

The 256 double symbols $E_\mu F_\nu$ ($\mu, \nu = 1, 2, \ldots 16$) constitute a complete set; that is to say, if a linear function of these symbols with algebraic coefficients is called an EF-number, the operations of addition, subtraction and multiplication applied to EF-numbers always yield EF-numbers. In matrix representation every EF-number is a fourfold matrix of the fourth rank, and conversely. A matrix of the fourth rank $T_{\alpha\gamma,\beta\delta}$ is resolved into components according to the formula

$$T_{\alpha\gamma,\beta\delta}=\Sigma_\mu\Sigma_\nu t_{\mu\nu}(E_\mu)_{\alpha\gamma}(E_\nu)_{\beta\delta} \qquad (10\cdot221)$$

or, without suffixes,

$$T=\sum_{\mu,\nu} t_{\mu\nu}E_\mu F_\nu. \qquad (10\cdot222)$$

The definition of pure wave tensors is extended to double wave tensors. A pure double tensor J is the product of two double wave vectors Ψ, X*, and we have

$$J=\Psi X^*=\Sigma j_{\mu\nu}E_\mu F_\nu. \qquad (10\cdot223)$$

A convention as to the order of suffixes in T or J is necessary. It is defined by (10·221). We write the product of the double vectors $\Psi_{\alpha\beta}$, $X_{\gamma\delta}{}^*$ as $J_{\alpha\gamma,\beta\delta}$. The comma separates "first" suffixes (in chain with E_μ) from "second" suffixes in chain with F_μ.

To determine $j_{\mu\nu}$, multiply both sides of (10·223) by $E_\sigma F_\tau$ and take the spur. The spur is formed by identifying the last suffix of each chain with the first suffix; thus the two chains contract separately. In particular we have spur $(E_\mu F_\nu)=$ spur $E_\mu \times$ spur F_ν. Hence, by (3·32),

$$\left.\begin{aligned}\text{spur } E_\mu F_\nu &= 0 \quad \text{if } \mu \text{ or } \nu \neq 16\\ &= -16 \text{ if } \mu=\nu=16.\end{aligned}\right\} \qquad (10\cdot231)$$

Hence

$$\text{spur}\,(\Sigma_{\mu,\nu}j_{\mu\nu}E_\mu F_\nu E_\sigma F_\tau)=\text{spur}\,(j_{\sigma\tau}E_\sigma{}^2F_\tau{}^2)=16j_{\sigma\tau}. \qquad (10\cdot232)$$

By (10·223) this can be written

$$\begin{aligned} 16j_{\sigma\tau} &= \text{spur}\,(\Psi\mathrm{X}^*E_\sigma F_\tau) \\ &= \text{spur}\,\{\Psi_{\alpha\beta}\mathrm{X}_{\gamma\delta}(E_\sigma)_{\gamma\epsilon}(F_\tau)_{\delta\zeta}\} \\ &= \Psi_{\alpha\beta}\mathrm{X}_{\gamma\delta}(E_\sigma)_{\gamma\alpha}(F_\tau)_{\delta\beta} \\ &= \mathrm{X}^*E_\sigma F_\tau \Psi. \end{aligned}$$

Thus the formula corresponding to (3·37) is

$$j_{\sigma\tau} = \tfrac{1}{16}\mathrm{X}^*E_\sigma F_\tau \Psi. \qquad (10{\cdot}24)$$

Let the symbols $E_\mu F_\nu$ in some conventional order be denoted by K_σ ($\sigma = 1, 2, \ldots 256$). Then $K_\sigma{}^2 = E_\mu{}^2 F_\nu{}^2 = 1$, and K_σ, K_τ either commute or anti-commute according to a fixed scheme. There are also multiplicative relations of the form $K_\sigma K_\tau = K_\lambda$ or iK_λ. These properties constitute the structure of the symbolic frame K_μ, and any other set of symbols with the same structure is an *equivalent* frame. Equivalent frames are obtained by the transformation

$$K_\mu' = qK_\mu q^{-1}, \qquad (10{\cdot}25)$$

where q is any non-singular K-number. For, as shown in § 2·7, this transformation does not change multiplicative relations.

The theory of double frames K_μ, double wave tensors $T_{\alpha\gamma,\beta\delta}$, double strain vectors $S_{\alpha\gamma,\beta\delta}$, etc., follows on similar lines to those of simple frames, tensors and vectors. Since q now contains 256 components, we have 256 independent relativity rotations, i.e. transformations without intrinsic change. The double space vectors are outer products of two complete space vectors, or sums of such products. We may call them complete space tensors of the second rank; but besides containing the ordinary components of a symmetrical tensor of the second rank, they include components of tensors of the third or fourth rank with some degree of antisymmetry (just as a complete space vector includes an antisymmetrical tensor of the second rank).

We notice that the new matrices K_μ' given by (10·25) are not in general simple products of E-matrices and F-matrices. For example, if $q = e^{\frac{1}{2}E_{12}F_2\theta}$,

$$\begin{aligned} E_1' &= e^{\frac{1}{2}E_{12}F_2\theta}E_1 e^{-\frac{1}{2}E_{12}F_2\theta} = E_1 e^{-E_{12}F_2\theta} \\ &= E_1\cos\theta + E_2F_2\sin\theta. \end{aligned}$$

This is evidently necessary because the K transformation has 256 coefficients, whereas separate transformations of E_μ and F_μ are defined by 32 coefficients.

Considering a system described by a double wave tensor T, the transformation $T' = qTq^{-1}$, where q is a K-number, will represent displacement without intrinsic change. For, as in § 2·9, the new system could have been obtained by carrying out the same construction in a different but equivalent frame $K' = qKq^{-1}$. This extension of the relativity principle is rather

difficult to grasp, because at first sight it seems to conflict with our experience. If we rotate the moon without rotating the earth, we create an observable difference in the angle between their axes, which must surely be counted as an intrinsic change of the double system. The earth and moon is scarcely a fair illustration; but it is tempting to think that there would be an analogous intrinsic change in a system of two elementary particles (protons or electrons), since these possess planes of spin. Now it is clear that, if there is anything in the double system which is intrinsically changed by the transformation, it is unrepresented in the symbolic tensor T.† It signifies therefore that the conceptual process of combining systems is something other than a pure multiplication; in other words it introduces an interaction. We have already had a hint as to the nature of this modification, namely that in a combined system we contemplate only "simultaneous" configurations of the particles. In this chapter we shall not consider interaction. Our point of view may be expressed by saying that we shall treat *double systems* as a preliminary to introducing the cementing interaction which will make them into *combined systems*.

It has to be remembered that particles described by simple stream vectors have an even probability distribution throughout spherical space. Relations between them are therefore not very comparable with the relations between localised objects such as the earth and moon. It is one of the features of interaction that it renders possible a more localised probability distribution, as is illustrated in the theory of the states of a hydrogen atom.

10·3. The Interchange Operator.

Since the complete sets E_μ, F_μ commute, they come under Case (b) of § 2·7. They are therefore connected by a transformation

$$F_\mu = PE_\mu P, \quad E_\mu = PF_\mu P, \quad P^2 = 1, \tag{10·31}$$

and

$$P = \tfrac{1}{4}\sum_1^{16} E_\mu F_\mu. \tag{10·32}$$

Since $P^2 = 1$, $\quad PE_\mu F_\nu P = PE_\mu P . PF_\nu P = F_\mu E_\nu$. (10·33)

And quite generally the operation $P(\ldots)P$ interchanges E and F within the bracket.

Let $\quad \gamma_{\mu\nu} = \tfrac{1}{2}(E_\mu F_\nu + E_\nu F_\mu), \quad \zeta_{\mu\nu} = \tfrac{1}{2}(E_\mu F_\nu - E_\nu F_\mu)$. (10·34)

Since $\quad P(E_\mu F_\nu \pm E_\nu F_\mu)P = E_\nu F_\mu \pm E_\mu F_\nu$

on multiplying by final P, we have

$$P(E_\mu F_\nu \pm E_\nu F_\mu) = (E_\nu F_\mu \pm E_\mu F_\nu)P,$$

so that

$$P\gamma_{\mu\nu} = \gamma_{\mu\nu} P, \quad P\zeta_{\mu\nu} = -\zeta_{\mu\nu} P. \tag{10·35}$$

† It may be expressed in the matrix representation of T, since matrices have the property of symmetry or antisymmetry which is not invariant for transformations. We have seen (§ 7·3) that this discrimination corresponds to laying down planes of simultaneity.

We can write, instead of (10·222),

$$T = T^s + T^a = \Sigma t_{\mu\nu}{}^s \gamma_{\mu\nu} + \Sigma t_{\mu\nu}{}^a \zeta_{\mu\nu}, \tag{10·361}$$

where
$$t_{\mu\nu}{}^s = t_{\mu\nu} + t_{\nu\mu}, \quad t_{\mu\nu}{}^a = t_{\mu\nu} - t_{\nu\mu}. \tag{10·362}$$

Then
$$PT^s = T^s P, \quad PT^a = -T^a P. \tag{10·363}$$

That is to say T is divided into symmetrical and antisymmetrical parts T^s, T^a, which respectively commute and anticommute with the interchange operator P.

There are 136 independent matrices $\gamma_{\mu\nu}$ in the symmetrical part, and 120 independent matrices $\zeta_{\mu\nu}$ in the antisymmetrical part.

Setting as usual $\bar{\Psi}_{\alpha\beta} = \Psi_{\beta\alpha}$, we can show that

$$P\Psi = -\bar{\Psi}, \quad \Psi^* P = -\bar{\Psi}^*. \tag{10·37}$$

By (10·31), $PE_\mu P\Psi = F_\mu \Psi$. That is to say, instead of applying the matrix operator F_μ to the second suffix of Ψ, we can first apply the change represented by P, apply the matrix operator to the first suffix, and then undo the change P (since $P^{-1} = P$). Evidently the operation P interchanges the two suffixes. It may include in addition a self-reciprocal operation Q which commutes with E_μ, so that $QE_\mu Q = E_\mu$. But since we may substitute any other E-symbol or F-symbol for E_μ in the foregoing argument, Q must commute with all the symbols and therefore be algebraic. Thus $Q = \pm 1$, since these are the only self-reciprocal algebraic operators.

To determine the sign, consider the special case $\Psi_{\alpha\beta} = \delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is the substitution operator. Then by (10·32)

$$\begin{aligned} P_{\alpha\gamma,\beta\delta}\delta_{\gamma\delta} &= \tfrac{1}{4}\Sigma (E_\mu)_{\alpha\gamma} (E_\mu)_{\beta\delta} \delta_{\gamma\delta} \\ &= \tfrac{1}{4}\Sigma (E_\mu)_{\alpha\gamma} (E_\mu)_{\beta\gamma} = \tfrac{1}{4}\Sigma (E_\mu \bar{E}_\mu)_{\alpha\beta}. \end{aligned}$$

Using four-point matrices, the ten symmetrical matrices give

$$E_\mu \bar{E}_\mu = E_\mu{}^2 = -1,$$

and the six antisymmetrical matrices give $E_\mu \bar{E}_\mu = 1$. Hence

$$\Sigma E_\mu \bar{E}_\mu = -10 + 6 = -4;$$

so that
$$P\delta = -1 = -\delta.$$

Accordingly the sign in (10·37) is verified.

The operator P can be factorised. Let

$$\left.\begin{aligned} P_1 &= \tfrac{1}{2}(E_{23}F_{23} + E_{31}F_{31} + E_{12}F_{12} - 1), \\ P_2 &= \tfrac{1}{2}(E_4 F_4 + E_5 F_5 + E_{45} F_{45} - 1). \end{aligned}\right\} \tag{10·38}$$

We easily find
$$-P_1 P_2 = \tfrac{1}{2}\sum_1^{16} E_\mu F_\mu = P.$$

Also
$$P_1{}^2 = 1, \quad P_2{}^2 = 1, \quad P_1 P_2 = P_2 P_1.$$

This factorisation is based on conjugate triads (§ 3·8), and the factors P_1, P_2 are the interchange operators for the two minor complete sets. The operator P_1 is commonly written†

$$P_1 = -\tfrac{1}{2}(\sigma_1\sigma_1' + \sigma_2\sigma_2' + \sigma_3\sigma_3' + 1), \qquad (10{\cdot}385)$$

a form obtained by writing $\sigma_\mu = i\zeta_\mu$ in (3·87). P_1 is then treated as the whole interchange operator, the factor P_2 being replaced by its eigenvalue 1. This is legitimate in the more elementary types of problem for which a single set of Pauli matrices suffices.

We can extend the theory of interchange to double sets. Let K_μ, L_μ ($\mu = 1, 2, \ldots 256$) be two double frames arranged in corresponding order, so that L_μ, L_ν commute or anticommute according as K_μ, K_ν commute or anticommute. By the method of §§ 2·7, 2·8, we can find an interchange operator P_{KL}, such that $L_\mu = P_{KL}K_\mu P_{KL}^{-1}$. If the L_μ are new symbols commuting with the K_μ, the operator is

$$P_{KL} = P_{KL}^{-1} = \tfrac{1}{16}\sum_1^{256} K_\mu L_\mu. \qquad (10{\cdot}39)$$

If, on the other hand, the L_μ are K-numbers, the factor $\frac{1}{16}$ is replaced by an adjustable algebraic coefficient α; also the singular case may arise, requiring us to substitute another reflection of K_μ.

10·4. Duality.

Consider a matrix of the fourth rank which is the product of four wave vectors‡

$$T_{\alpha\gamma,\beta\delta} = \psi_\alpha\phi_\beta\chi_\gamma\omega_\delta. \qquad (10{\cdot}411)$$

Since α and γ are "first" suffixes, ψ and χ are wave vectors in the frame E_μ; ϕ and ω are wave vectors in the frame F_μ. Let

$$\Psi = \psi\omega^*, \quad \mathrm{X} = \phi\chi^*, \quad U = \psi\chi^*, \quad V = \phi\omega^*. \qquad (10{\cdot}412)$$

Then U and V are tensors of the second rank in the frames E_μ and F_μ respectively; Ψ and X, being composed of one vector from each frame, are double wave vectors. We have

$$T_{\alpha\gamma,\beta\delta} = U_{\alpha\gamma}V_{\beta\delta} = \Psi_{\alpha\delta}\mathrm{X}_{\beta\gamma}. \qquad (10{\cdot}413)$$

We can associate Ψ and X with new frames C_μ' and D_μ' in the same way that U and V are associated with the frames E_μ and F_μ.§ When the "crossed frames" C_μ' and D_μ' are adopted, Ψ and X become wave tensors of the second rank and U and V become double wave vectors.

This gives two alternative resolutions of T into matrix components, viz.

$$T = \Sigma t_{\mu\nu}E_\mu F_\nu = \Sigma\tau_{\mu\nu}'C_\mu'D_\nu' \qquad (10{\cdot}421)$$

† Dirac, *Quantum Mechanics*, 2nd ed., p. 226, equation (35). Dirac uses Pauli matrices whose square is +1.

‡ It is not necessary that the factorisation should be actually possible; the theory in this section will apply to symbolic factors (§ 3·3).

§ Unaccented frames C, D are introduced later.

or, with suffixes,

$$T_{\alpha\gamma,\beta\delta}=\Sigma t_{\mu\nu}(E_\mu)_{\alpha\gamma}(E_\nu)_{\beta\delta}=\Sigma \tau_{\mu\nu}'(E_\mu)_{\alpha\delta}(E_\nu)_{\beta\gamma}. \quad (10{\cdot}422)$$

The distinctive notation C', D', E, F is not required when suffixes are inserted.

Evidently a double matrix $E_\mu F_\nu$ is a linear function of the 256 symbols $C_\mu' D_\nu'$, and *vice versa*. We shall determine this function explicitly. Let

$$E_\mu F_\nu=\Sigma_{\sigma,\tau} v'_{\mu\nu,\sigma\tau} C_\sigma' D_\tau' \quad (10{\cdot}431)$$

or, with suffixes, $(E_\mu)_{\alpha\gamma}(E_\nu)_{\beta\delta}=\Sigma_{\sigma,\tau} v'_{\mu\nu,\sigma\tau}(E_\sigma)_{\alpha\delta}(E_\tau)_{\beta\gamma}$. (10·432)

Multiply both sides by $(E_\lambda)_{\delta\alpha}(E_\rho)_{\gamma\beta}$. The right-hand side is

$$\Sigma_{\sigma,\tau} v'_{\mu\nu,\sigma\tau}\,\mathrm{spur}\,(E_\sigma E_\lambda).\mathrm{spur}\,(E_\tau E_\rho)=16v'_{\mu\nu,\lambda\rho}.$$

The left-hand side is

$$(E_\mu)_{\alpha\gamma}(E_\rho)_{\gamma\beta}(E_\nu)_{\beta\delta}(E_\lambda)_{\delta\alpha}=\mathrm{spur}\,(E_\mu E_\rho E_\nu E_\lambda).$$

Hence

$$v'_{\mu\nu,\lambda\rho}=\tfrac{1}{16}\,\mathrm{spur}\,(E_\mu E_\rho E_\nu E_\lambda). \quad (10{\cdot}44)$$

As an important example, consider $E_{16}F_{16}$. Let

$$E_{16}F_{16}=\Sigma_{\sigma,\tau} v'_{\sigma\tau} C_\sigma' D_\tau' \quad (10{\cdot}451)$$

or, with suffixes, $-(1)_{\alpha\gamma}(1)_{\beta\delta}=\Sigma_{\sigma,\tau} v'_{\sigma\tau}(E_\sigma)_{\alpha\delta}(E_\tau)_{\beta\gamma}$. (10·452)

Then, by (10·44), $v'_{\sigma\tau}=-\tfrac{1}{16}\,\mathrm{spur}\,(E_\tau E_\sigma)$ (10·46)

so that $v'_{\sigma\sigma}=\tfrac{1}{4}$, $v'_{\sigma\tau}=0$ $(\sigma\neq\tau)$. Hence

$$E_{16}F_{16}=\tfrac{1}{4}\Sigma_\sigma C_\sigma' D_\sigma'=P_{CD}', \quad (10{\cdot}471)$$

where P_{CD}' is the interchange operator of the frames C_μ' and D_μ'. Similarly

$$C_{16}'D_{16}'=P_{EF}. \quad (10{\cdot}472)$$

Thus by crossing the frames, algebraic quantities in the frame $C'D'$ become multiples of the interchange operator of the frame EF, and *vice versa*.

We find similarly that $E_\mu F_\mu=\tfrac{1}{4}\Sigma_\sigma(\pm C_\sigma' D_\sigma')$, (10·48)

where the minus sign is to be taken if E_σ anticommutes with E_μ.

A slight variation of the foregoing process is obtained by taking frames C_μ, D_μ such that the connection

$$T=\Sigma t_{\mu\nu} E_\mu F_\nu=\Sigma \tau_{\mu\nu} C_\mu D_\nu \quad (10{\cdot}491)$$

stands for $T_{\alpha\gamma,\beta\delta}=\Sigma t_{\mu\nu}(E_\mu)_{\alpha\gamma}(E_\nu)_{\beta\delta}=\Sigma \tau_{\mu\nu}(E_\mu)_{\alpha\beta}(E_\nu)_{\gamma\delta}$. (10·492)

This "straight cross" is easier to manipulate analytically, and we shall generally use it; though it is perhaps less physically significant than the "inverted cross". If $E_\mu F_\nu=\Sigma v_{\mu\nu,\sigma\tau} C_\sigma D_\tau$

we obtain, by the same method as before,

$$v_{\mu\nu,\sigma\tau}=\tfrac{1}{16}\,\mathrm{spur}\,(E_\sigma E_\mu \bar{E}_\tau \bar{E}_\nu), \quad (10{\cdot}493)$$

where, as usual, $(\bar{E}_\tau)_{\alpha\beta}=(E_\tau)_{\beta\alpha}$. Hence

$$E_{16}F_{16}=\tfrac{1}{4}\Sigma(\pm C_\sigma D_\sigma), \tag{10·494}$$

the minus sign applying to the six antisymmetrical (time-like) matrices. Also

$$E_{45}F_{45}=\tfrac{1}{4}\Sigma(\pm C_\sigma D_\sigma), \tag{10·495}$$

the minus sign applying to the six electrical matrices $\sigma=01, 02, 03, 04, 05, 16$. The same formulae hold if EF and CD are interchanged.

We call the tensor $\tau_{\mu\nu}$ the *dual* of the tensor $t_{\mu\nu}$.

Denoting the double matrices $E_\mu F_\nu$, $C_\mu D_\nu$ by K_σ, L_σ ($\sigma=1, 2, \ldots 256$), the equivalent frames K_σ and L_σ are connected by a transformation

$$L_\sigma=qK_\sigma q^{-1}. \tag{10·496}$$

It should not be difficult to evaluate q as an EF-number or CD-number; but I have not succeeded in doing so. The calculation is complicated because it is found to come under the singular case; the value of P_{KL} calculated from (10·39) is zero, so that $q\neq P_{KL}$. The same applies to the transformation connecting $E_\mu F_\nu$ and $C_\mu' D_\nu'$.

Without actually evaluating q, we see that the crossing of frames is included in the relativity transformations of the double frame. We can regard the transformation as a continuous one, giving a sequence of double frames intermediate between EF and CD.

10·5. Significance of the Crossed Frames.

According to the uncertainty principle there are two extreme ways of specifying a particle: (*a*) the momentum factor may be specified exactly, so that the position vector is entirely uncertain, and (*b*) the position vector may be specified exactly, so that the momentum vector is entirely uncertain. We shall find that the effect of crossing frames is to transform specification (*a*) into specification (*b*). That is to say, a particle which has the specification (*a*) in the EF frame has the specification (*b*) in the CD frame, and *vice versa*.

Double frames enable us to express symmetrical space tensors of the second rank in wave tensor form. (A single frame provides only for the antisymmetrical tensors of the second rank.) Consider, for example, the outer square $p_\mu p_\nu$ of a momentum vector p_μ. We call $p_\mu p_\nu$ the energy tensor,† or more precisely the *self-energy tensor* of the particle. In ordinary relativity theory the energy tensor of a particle is $p_\mu p_\nu/m$; but the factor m belongs to a later stage in the adaptation of our equations to practical conditions, and can be omitted for the present. By (6·51) the four-dimensional momentum vector is represented by $\Sigma E_{\mu 5}p_\mu$; and the energy tensor is therefore represented by

$$\Sigma E_{\mu 5}F_{\nu 5}p_\mu p_\nu \quad (\mu,\nu=1,2,3,4). \tag{10·51}$$

† Both the momentum vector and the energy tensor include momentum and energy. It is an accident of current nomenclature that they are named differently.

If the momentum vector is exact, we may take its direction as our time axis. The energy tensor (10·51) then reduces to a single component

$$\alpha E_{45} F_{45}. \tag{10·521}$$

To represent an entirely uncertain momentum we must distribute the probability evenly over momentum vectors in all directions in space-time. The resultant momentum will be zero, but the resultant energy tensor will be definite. "Entire uncertainty" implies that there is no preference for any direction in space-time; the resultant energy tensor must therefore be of a form invariant for rotations and Lorentz transformations. The only invariant tensor of the second rank is $\delta_\mu{}^\nu$. Hence

$$\Sigma p_\mu p_\nu = \beta \delta_\mu{}^\nu,$$

where β is an invariant, and the summation extends over all the vectors which compose the probability distribution. Then by (10·51) the resultant energy tensor is

$$\beta (E_{15} F_{15} + E_{25} F_{25} + E_{35} F_{35} + E_{45} F_{45}). \tag{10·522}$$

To extend this from simple space vectors to complete space vectors, we define the *complete self-energy tensor* to be the outer square of the complete stream vector. If the complete stream vector is exact, it can be reduced to the form $\alpha^{\frac{1}{2}} (E_1 + E_{23} + E_{45} + E_{16})$, and the complete energy tensor is

$$\alpha (E_1 + E_{23} + E_{45} + E_{16}) (F_1 + F_{23} + F_{45} + F_{16}). \tag{10·523}$$

This refers to an elementary charged particle; for a neutral particle the complete tensor is still of the form (10·521).

If the stream vector is entirely uncertain, the distribution must be such that the resultant energy tensor is invariant for all appropriate relativity rotations. The natural generalisation of (10·522) is

$$\beta \sum_1^{16} E_\mu F_\mu. \tag{10·524}$$

But this corresponds to a probability distribution which is symmetrical with respect to an electrically saturated space-time. To obtain symmetry with respect to neutral space-time, we must substitute neutral sets (6·412). Distinguishing kinematical and electrical matrices by suffixes k and e, the resultant complete energy tensor is then

$$\beta (\Sigma_k E_\mu F_\mu - \Sigma_e E_\mu F_\mu) = 4\beta C_{45} D_{45} \tag{10·525}$$

by (10·495). Comparing with (10·521) we see that this corresponds to a neutral particle† with definite stream vector in the CD frame.

This establishes the important result that for a neutral particle an entirely uncertain distribution of momentum in the EF frame corresponds to an exact momentum in the CD frame. (The momentum is here understood to include energy, spin and magnetic moment.)

† When the complete uncertainty extends to all components of the stream vector, the charge must be completely uncertain.

We see also that the straight cross CD is involved because we take a neutral frame as standard. The theoretically simpler inverted cross $C'D'$ corresponds to a saturated EF frame.

The position vector can be treated similarly. We can define a *complete position vector* which includes, for example, the angular coordinates conjugate to the spin momenta. For a neutral particle it reduces to the ordinary four-dimensional position vector, since there is no spin or magnetic moment to be described. The outer square of the position vector will be called the *position tensor*; it plays an analogous part to the energy tensor in the foregoing theory.

The energy tensor and the position tensor have definite values even when the corresponding vectors are uncertain; but for brevity we shall call those of type (10·521) "exact", and those of type (10·525) "uncertain". If a particle is represented by an exact energy tensor and an uncertain position tensor in the EF frame, it is represented by an uncertain energy tensor and an exact position tensor in the CD frame. If the scale relation is appropriately chosen, the exact energy tensor of the EF frame is identical with the exact position tensor of the CD frame; and similarly the two uncertain tensors are identical. In short the position space of the EF frame is the momentum space of the CD frame and *vice versa*.

We have here adopted the four-dimensional point of view, in which the position vector is measured from an origin in space-time. This was necessary in order to compare our formulae with the ordinary form of the uncertainty principle; but it slurs over the effect of curvature of space-time.† In five-dimensional representation the position vector is measured from an origin at the centre of curvature of space-time; and in the standard notation an exact position tensor has the form

$$\alpha E_5 F_5 = \tfrac{1}{4}\alpha\Sigma(\pm C_\mu D_\mu), \tag{10·53}$$

where the minus sign refers to $\mu = 04, 14, 24, 34, 54, 16$, by applying (10·493). The connection between the position space and the momentum space is, as it were, rotated by the matrix E_4. The connection is expressed more simply by correlating the four-dimensional density of the position vector to the three-dimensional density of the momentum vector. Each of these gives a second-rank tensor of the form‡

$$\alpha E_{16} F_{16} = \tfrac{1}{4}\alpha(\Sigma_s C_\mu D_\mu - \Sigma_t C_\mu D_\mu), \tag{10·54}$$

† If the position were entirely uncertain in infinite space, the position tensor would be infinite. In assuming it to be finite, we have in a sense taken account of curvature of space-time; but it has not been necessary to introduce the curvature in any other way.

‡ Multiplying the position vector (matrix E_5) by iE_5 to obtain its four-dimensional density, and multiplying the momentum vector (matrix E_{45}) by iE_{45} to obtain its three-dimensional density, the result is in each case an algebraic quantity (matrix E_{16}); so that the corresponding second rank tensor is $\alpha E_{16} F_{16}$.

where s and t refer to the space-like and time-like matrices. It is these densities which become interchanged in the transformation $EF \to CD$.

Formally an energy tensor $\Sigma E_\mu F_\nu p_\mu p_\nu$ refers to two particles or systems, since it introduces two frames. It is a special case of a *mutual energy tensor* of two particles $\Sigma E_\mu F_\nu p_\mu p_\nu'$, formed by multiplication of their stream vectors $\Sigma E_\mu p_\mu$, $\Sigma F_\nu p_\nu'$, as the combination of probabilities requires. A *self-energy tensor* is obtained by imposing the constraint $p_\mu = p_\mu'$ (or alternatively $p_\mu = -p_\mu'$); this constraint greatly limits the number of relativity transformations applicable. There are two possible interpretations of such a constraint; either the second particle is a fictitious duplication of the first, introduced for formal purposes, or the centroid of the two particles is deemed to be fixed so that their stream vectors are automatically made equal and opposite.

This study of a double frame throws considerable light on the relation of position and momentum in the matrix theory. But it is only a step towards the actual conditions in which position and momentum are known observationally; and further progress will be made in Chapter XI.

10·6. The 136-dimensional Phase Space.

When four-point matrices are used, the frames E_μ, F_μ each contain ten imaginary and six real matrices. Hence the frame $E_\mu F_\nu$ contains

$$(10 \times 10) + (6 \times 6) = 136 \text{ real matrices}$$

and

$$(6 \times 10) + (10 \times 6) = 120 \text{ imaginary matrices.}$$

Since $(E_\mu F_\nu)^2 = 1$, the eigenvalues of $E_\mu F_\nu$ are real. Accordingly, by the definition in § 7·3, the 136 real matrices are space-like and the 120 imaginary matrices are time-like.

The general infinitesimal transformations for covariant final and contravariant initial double wave vectors are

$$\Psi' = e^{\frac{1}{2}d\Theta}\Psi, \quad X^{*\prime} = X^* e^{-\frac{1}{2}d\Theta}, \tag{10·61}$$

where $d\Theta = \Sigma E_\mu F_\nu \theta_{\mu\nu}$. Considering a single term $\theta_{\mu\nu}$, the two transformations in (10·61) are the same if

$$\exp\{\tfrac{1}{2}(E_\mu)_{\alpha\gamma}(E_\nu)_{\beta\delta}\theta_{\mu\nu}\} \cdot \Psi_{\gamma\delta} = \Psi_{\gamma\delta} \cdot \exp\{-\tfrac{1}{2}(E_\mu)_{\gamma\alpha}(E_\nu)_{\delta\beta}\theta_{\mu\nu}\}.$$

That is, if

$$-(E_\mu)_{\gamma\alpha}(E_\nu)_{\delta\beta} = (E_\mu)_{\alpha\gamma}(E_\nu)_{\beta\delta}$$

or

$$-\bar{E}_\mu \bar{F}_\nu = E_\mu F_\nu. \tag{10·62}$$

This requires that $E_\mu F_\nu$ should be imaginary or time-like. Thus when $E_\mu F_\nu$ is time-like, a covariant wave vector transforms like a contravariant wave vector, and a covariant wave tensor transforms like a mixed wave tensor.

Proceeding as in § 7·2 we obtain double strain vectors (covariant double wave tensors), which behave like double space vectors for time-like transformations and substitute antiperpendicular rotations for relativity

rotations in space-like transformations. The latter generate a 136-dimensional phase space.

The theory of ten-dimensional phase space applies without important modification to 136 dimensions. Stereographic coordinates can be introduced; and the result corresponding to (7·59) is

$$\sqrt{-g} = R^{-136}(1+r^2/4R^2)^{-136}. \tag{10·63}$$

Since $(E_\mu F_\nu)^2 = 1$, the circular rotations are of the form $e^{iE_\mu F_\nu u_{\mu\nu}}$, where $u_{\mu\nu}$ is real. Thus the matrix $d\Theta_s$ for real displacements in phase space is imaginary, as in simple phase space (7·31); but the coefficients $\theta_{\mu\nu}$ (unlike the θ_μ) are imaginary.

The condition $\bar{E}_\mu \bar{F}_\nu = E_\mu F_\nu$ satisfied by the space-like matrices means that they are symmetrical for an interchange of the suffixes α, β with γ, δ in (10·492). There is another kind of symmetry which corresponds to the interchange of α, γ with β, δ, i.e. interchange of first suffixes with second suffixes. The double matrices $\gamma_{\mu\nu}$ introduced in § 10·3 have this kind of symmetry, and the $\zeta_{\mu\nu}$ have the corresponding antisymmetry. It is inconvenient to work with $\gamma_{\mu\nu}$ and $\zeta_{\mu\nu}$, since they have no simple commutative properties and their squares are not algebraic. But the two kinds of symmetry are interchanged by crossing the frames as in (10·492); hence the space-like and time-like matrices of the *CD* frame have the same symmetry and antisymmetry as the $\gamma_{\mu\nu}$ and $\zeta_{\mu\nu}$. It is easily seen that the 136 $\gamma_{\mu\nu}$ are linear functions of the 136 space-like matrices of the *CD* frame and the 120 $\zeta_{\mu\nu}$ are linear functions of the 120 time-like matrices.

Thus the separation of T into $T^s + T^a$ in the *EF* frame (10·361) is the same as the separation into space-like and time-like parts in the *CD* frame.

For any double system we have two alternative phase spaces according as we adopt the *EF* frame or the *CD* frame. If one of them corresponds to classification of configuration by position, the other will correspond to classification by momentum. If a simple *E* or *F* phase space gives a classification by position, the *EF* phase space will give the position configurations, and the *CD* phase space will give the momentum configurations. For example, an algebraic displacement in *EF* space represents change of time, and in the *CD* space represents change of energy. Since the points of *CD* space represent distributions with fixed momentum, generally called elementary states, we may regard the *CD* space as composed of *states* and the *EF* space as composed of *configurations*.

It is perhaps not self-evident that the *EF* space gives the same kind of classification as the simple *E* and *F* spaces.† For a crucial test consider an electrical displacement with matrix E_1 for the space vector or E_{23} for the strain vector. To give this displacement to both particles we must apply

† It had occurred to me that the process of amalgamating two simple systems to form a combined system might involve crossing their frames. The test here given dispels this idea.

the transformation $q = e^{\frac{1}{2}(E_{23}+F_{23})\theta}$ to the double strain vector. Since this is a $\gamma_{23,16}$ rotation it gives displacement in CD space; there is no displacement in EF phase space (or in the separate E and F phase spaces) since the matrices E_{23}, F_{23} (or $E_{23}F_{16}$, $E_{16}F_{23}$) are time-like. It can scarcely be supposed that we create a *positional* displacement of the double system without displacing the simple systems positionally; but the displacement in CD space can well be interpreted as a change of electrical *energy*, which has no counterpart in a simple system. This indicates that the EF phase space in which no change occurs is the positional space, and the CD phase space is the momentum space.

The following property of CD phase space is of interest. If we divide a double wave vector into symmetrical and antisymmetrical parts Ψ^s, Ψ^a, so that $\Psi^s = \bar{\Psi}^s$, $\Psi^a = -\bar{\Psi}^a$, we have, by (10·37),

$$\Psi^s = \tfrac{1}{2}(1-P)\Psi, \quad \Psi^a = \tfrac{1}{2}(1+P)\Psi. \tag{10·64}$$

Consider a transformation $\Psi' = q\Psi$. If q commutes with P, we obtain

$$(\Psi')^s = q\Psi^s, \quad (\Psi')^a = q\Psi^a. \tag{10·65}$$

But for a non-commuting q, say $q = 1 + \zeta_{\mu\nu}\theta$,

$$\begin{aligned}(\Psi')^s &= \tfrac{1}{2}(1-P)(1+\zeta_{\mu\nu}\theta)\Psi \\ &= \Psi^s + \tfrac{1}{2}\zeta_{\mu\nu}\theta(1+P)\Psi \\ &= \Psi^s + \zeta_{\mu\nu}\theta\Psi^a.\end{aligned} \tag{10·66}$$

Thus Ψ^s and Ψ^a are kept distinct in the $\gamma_{\mu\nu}$ transformations but not in the $\zeta_{\mu\nu}$ transformations. The former comprise all the displacements in CD phase space. Thus CD phase space has the distinctive property that a strain vector of the form $\Psi^s\Phi^s$ remains of the same form at all points of phase space; similarly for the forms $\Psi^a\Phi^a$, $\Psi^s\Phi^a$. Thus states determined by symmetrical and antisymmetrical wave vectors Ψ^s, Ψ^a form distinct systems—not transformable into one another by strain.

This separation of symmetrical and antisymmetrical wave functions plays a prominent part in the current treatment of double systems, as developed by Fermi and Dirac. Since it applies to CD space but not to EF space, it applies to states not to configurations. Since this is in accordance with the current interpretation, it confirms our identification of the two phase spaces.

10·7. Representations of Two Charges.

The strain vectors representing elementary charged particles have been given in (6·64). Using the matrix representation in (3·61), we resolve them into their wave vector factors

$$\begin{aligned}S_a &= m(1, 1, -i, -i)(1, 1, i, i), \\ S_b &= m(1, 1, i, i)(1, 1, -i, -i), \\ S_c &= m(1, -1, -i, i)(1, -1, i, -i), \\ S_d &= m(1, -1, i, -i)(1, -1, -i, i).\end{aligned} \tag{10·71}$$

First consider two charges of opposite sign and the same spin, whose wave tensors may be taken to be

$$U=\psi\chi^*=S_a, \quad V=\phi\omega^*=S_b,$$

so that (omitting the masses m, m') we have

$$\psi=\omega=(1,\ 1,\ -i,\ -i), \quad \chi=\phi=(1,\ 1,\ i,\ i).$$

The double vectors, obtained by a straight cross, are therefore

$$\Psi=\psi\phi^*=S_a, \quad \mathrm{X}=\chi\omega^*=S_b. \tag{10·72}$$

If the masses are inserted, Ψ and X can be given any coefficients m'', m''' such that $m''m'''=mm'$.

As another example take two charges of the same sign and spin, with $U=V=S_a$, so that

$$\psi=\phi=(1,\ 1,\ -i,\ -i), \quad \chi=\omega=(1,\ 1,\ i,\ i).$$

Then

$$\begin{aligned}\Psi=\psi\phi^*&=(1,\ 1,\ -i,\ -i)(1,\ 1,\ -i,\ -i)\\ &=\begin{matrix}1 & 1 & -i & -i\\ 1 & 1 & -i & -i\\ -i & -i & -1 & -1\\ -i & -i & -1 & -1\end{matrix}\\ &=-i(C_{35}+C_{42}-iC_{43}-iC_{52}),\end{aligned} \tag{10·73}$$

so that the double wave vectors are degenerate (5·56).

A combination of two charges of like sign and opposite spin, or of opposite sign and spin, likewise gives degenerate double wave vectors.

We see that there is a simplicity in the combination of charges of opposite sign with like spin, which is not exhibited by any other combination. The CD and the EF representations are interchangeable, and the double wave vectors are non-degenerate. The physical interpretation of this peculiarity is that when charges of opposite sign with like spin are superposed, their electric and magnetic fields cancel. They thus constitute a self-contained unit which can be inserted in a background of neutral space-time without disturbing its neutrality. Other combinations cannot be treated as isolated units for they disturb by their external fields the surrounding matter. In their case we must take into account, in calculating probabilities for the purposes of statistical theory, not only the probability of the configuration of the two particles, but the probability of the polarised configuration of the rest of the universe.

If then we wish to treat what is strictly a two-body problem, we must confine ourselves to particles of opposite sign and like spin.

To understand the nature of the *superposition* referred to above, we must recall that the momentum vectors of the two charges are given definitely

by S_a and S_b. Hence, by the uncertainty principle, their positions are entirely indeterminate. We are therefore superposing two probability distributions which extend uniformly over the universe. It is not a question of superposing localised particles; they would be disturbed by interaction effects which we have not yet studied; and their exact superposition is, in fact, inhibited. The theory of this chapter is limited to the combination of independent probabilities.

It will be seen from (10·71), or directly from (6·64), that

$$S_b = \bar{S}_a, \quad S_d = \bar{S}_c. \tag{10·74}$$

By making an inverted cross we obtain double vectors $\Psi'' = \psi\omega^*$, $X' = \phi\chi^*$. These are non-degenerate if the two charges are of like sign and spin, and are degenerate for all other combinations. We have already noticed that the inverted cross bears the same relation to electrically saturated space that the straight cross bears to neutral space; and the non-degenerate combination of two charges of like sign and spin would evidently fit saturated space in the same way that two charges of opposite sign and like spin fit neutral space.

10·75. Importance of Double Tensors.

The fundamental tensors of relativity theory are symmetrical space tensors of the second rank, viz. the metrical tensor $g_{\mu\nu}$ and the energy tensor $G_{\mu\nu} - \frac{1}{2}g_{\mu\nu}G$. These must appear as double wave tensors in the wave-tensor calculus; and it is clear that any serious attempt to unify relativity theory with wave mechanics must be based on double wave tensors. In macroscopic relativity theory the interval ds between two points is regarded as the fundamental observable; correspondingly in relativistic wave mechanics the double wave tensor of two particles is the basis of description.

In treating a system of n particles, we naturally introduce a corresponding n-tuple wave tensor. But the step from a double to an n-tuple wave tensor is not comparable with the step from a single to a double wave tensor. In relativity theory we do not recognise any relation between three points analogous to the interval between two points; nor are there any space tensors of the nth rank analogous to the metrical tensor and the energy tensor. In proceeding from simple to double wave tensors we approach the meeting point with ordinary relativity theory; in proceeding to treble and n-tuple wave tensors we diverge again. Quantum physicists have themselves recognised that the specification of a system of n particles cannot well be left in the form of an n-tuple wave tensor, and have introduced another method of specifying such systems by a sum (not product) of Jordan-Wigner wave functions, which will be used in Chapter XVI.

The number 136, being the number of dimensions of double phase space, will be prominent in our formulae for the principal constants of nature. This

does not imply that their application has special reference to systems consisting of two particles. It implies that in wave mechanics, as in relativity theory, the study of a complex system is based on the relations between pairs of particles contained in it, and on the tensors of the second rank which embody these relations. The number of dimensions of double phase space occurs in the formula for the mass of an electron or proton, because mass is by definition contained in a space tensor of the second rank equivalent to a double wave tensor.

The quadruple tensor also plays a rather special part in the theory. If a primitive entity or *relatum* is represented by a simple wave tensor, we require two relata in order to provide an observable *relation*. To assign measure to this relation we require another observable relation with which to compare it. Thus a *measure* involves four relata—two to provide a quantity to be measured and two to provide a unit of comparison. It therefore depends on a quadruple wave tensor. This coincides with our conclusion in general relativity theory:†

> Thus four points is the minimum number for which an assertion of absolute structural relation can be made. The ultimate elements of structure are thus four-point elements.

Thus simple, double and quadruple wave tensors appear in the theory; but there is no occasion for introducing any other combination for the purposes of fundamental study. For systems of more than four particles (or two particles plus a "comparison fluid") we abandon the method of multiplication of probabilities, and develop a procedure (formally investigated in Chapter XVI) which is in a sense intermediate between macroscopic theory and elementary wave mechanics.

Generally the quadruple wave tensor remains in the background of the theory; we need only occasionally remind ourselves that the double wave tensors, which we employ, require such a background. We shall generally regard the quantities described by double wave tensors as observables, in the same way that the interval ds is usually regarded as an observable although any numerical value that is attached to it expresses its relation to another interval.

10·8. Relative Coordinates.

Let x_μ, x_μ' be the coordinates of two points in four dimensions, associated respectively with symbolic frames E_μ, F_μ. We make the transformation

$$x_\mu = \tfrac{1}{2}(x_\mu + x_\mu'), \quad \xi_\mu = \tfrac{1}{2}(x_\mu - x_\mu'), \tag{10·811}$$

which gives

$$x_\mu = x_\mu + \xi_\mu, \quad x_\mu' = x_\mu - \xi_\mu. \tag{10·812}$$

† *Mathematical Theory of Relativity*, § 98.

Then the x_μ are coordinates of the centroid, and $\pm 2\xi_\mu$ are the coordinates of one point relative to the other. The corresponding formulae for the derivatives are

$$\frac{\partial}{\partial x_\mu} = \frac{\partial}{\partial x_\mu} + \frac{\partial}{\partial x_\mu'}, \quad \frac{\partial}{\partial \xi_\mu} = \frac{\partial}{\partial x_\mu} - \frac{\partial}{\partial x_\mu'}, \tag{10·821}$$

$$\frac{\partial}{\partial x_\mu} = \frac{1}{2}\left(\frac{\partial}{\partial x_\mu} + \frac{\partial}{\partial \xi_\mu}\right), \quad \frac{\partial}{\partial x_\mu'} = \frac{1}{2}\left(\frac{\partial}{\partial x_\mu} - \frac{\partial}{\partial \xi_\mu}\right). \tag{10·822}$$

We are going to find two new complete sets G_μ, H_μ which, we shall show, are associated with x_μ, ξ_μ in the same way that E_μ, F_μ are associated with x_μ, x_μ'; so that H_μ is the frame whose rotations determine the vector transformations of the relative coordinates $2\xi_\mu$.

Let P be the interchange operator (§ 10·3) of E_μ and F_μ, and let

$$G_\mu = \frac{P}{1+i'}(E_\mu + i'F_\mu), \quad H_\mu = \frac{P}{1+i'}(F_\mu + i'E_\mu), \tag{10·831}$$

for $\mu = 1, 2, 3, 4$. We here write i' for an algebraic square root of -1, not necessarily the same square root as E_{16} and F_{16}. Solving (10·831) for E_μ, F_μ we find

$$E_\mu = \frac{P}{1-i'}(G_\mu - i'H_\mu), \quad F_\mu = \frac{P}{1-i'}(H_\mu - i'G_\mu). \tag{10·832}$$

Then

$$G_\mu^2 = (1/2i')\,P\,(E_\mu + i'F_\mu)\,P\,(E_\mu + i'F_\mu) = (1/2i')\,(F_\mu + i'E_\mu)\,(E_\mu + i'F_\mu)$$
$$= -1.$$

Similarly
$$G_\mu G_\nu = (1/2i')\,(F_\mu + i'E_\mu)\,(E_\nu + i'F_\nu)$$
$$= (1/2i')\,\{F_\mu E_\nu - F_\nu E_\mu + i'\,(E_\mu E_\nu + F_\mu F_\nu)\}. \tag{10·841}$$

Since μ, ν are restricted to the values 1, 2, 3, 4, the right-hand side is antisymmetrical in μ and ν. Therefore $G_\mu G_\nu = -G_\nu G_\mu$. Similarly $H_\mu H_\nu = -H_\nu H_\mu$. Also

$$G_\mu H_\nu = (1/2i')\,(F_\mu + i'E_\mu)\,(F_\nu + i'E_\nu), \quad H_\nu G_\mu = (1/2i')\,(E_\nu + i'F_\nu)\,(E_\mu + i'F_\mu). \tag{10·842}$$

These two expressions are equal, so that $G_\mu H_\nu = H_\nu G_\mu$.

These results show that G_1, G_2, G_3, G_4 and H_1, H_2, H_3, H_4 are tetrads, and that the H's commute with the G's. From these tetrads we construct two complete sets G_μ, H_μ ($\mu = 1, 2, \ldots 16$) which commute with one another. By (10·842)

$$G_\mu H_\mu = (1/2i')\,(F_\mu + i'E_\mu)^2 = E_\mu F_\mu \tag{10·843}$$

for $\mu = 1, 2, 3, 4$. This result can be extended to all values of μ; e.g.

$$G_{12}H_{12} = G_1G_2H_1H_2 = G_1H_1 . G_2H_2 = E_1F_1 . E_2F_2 = E_{12}F_{12}.$$

Hence by (10·32) the interchange operator P for E_μ and F_μ is also the interchange operator for G_μ and H_μ.

The G_μ and H_μ not belonging to the original tetrad do not obey (10·831). In general they have no simple connection with the corresponding E_μ and F_μ; but the result for G_5, H_5 is of interest. We have

$$\begin{aligned} G_5 &= -iG_1G_2G_3G_4 \\ &= \tfrac{1}{4}iP(E_1+i'F_1)P(E_2+i'F_2)P(E_3+i'F_3)P(E_4+i'F_4) \\ &= \tfrac{1}{4}i(i'E_1+F_1)(E_2+i'F_2)(i'E_3+F_3)(E_4+i'F_4) \qquad (10\cdot85) \\ &= -\tfrac{1}{4}iE_1E_2E_3E_4(1-i'E_4F_4-E_{34}F_{34}-i'E_{15}F_{15}-E_5F_5-\ldots). \qquad (10\cdot861) \end{aligned}$$

The bracket contains 16 terms. We have given one specimen of each type of term. All terms of the same type have the same coefficient; the verification of this is an interesting study of the operation of the permutations. We obtain similarly from (10·85)

$$G_5 = -\tfrac{1}{4}iF_1F_2F_3F_4(1+i'E_4F_4-E_{34}F_{34}+i'E_{15}F_{15}-E_5F_5+\ldots). \qquad (10\cdot862)$$

Multiplying (10·861) by E_5 and (10·862) by $i'F_5$ and adding, we obtain

$$(E_5+i'F_5)G_5 = (1+i')P.$$

Hence, multiplying initially by P and finally by G_5,

$$G_5 = -\frac{P}{1+i'}(E_5+i'F_5) = -G_5', \qquad (10\cdot863)$$

where G_5' is the symbol defined analogously to G_1, G_2, G_3, G_4 by putting $\mu=5$ in (10·831). A similar result is obtained for H_5.

Thus we can, if we wish, extend the present investigation to relative coordinates in five dimensions, defining G_μ, H_μ ($\mu=1, 2, 3, 4, 5$) by (10·831); but in that case if E_μ, F_μ are right-handed sets we must construct G_μ, H_μ as left-handed sets. Since we have at present no use for this extension, we keep to right-handed sets, and accordingly define G_5 as $-iG_1G_2G_3G_4$.

By (10·811) and (10·831) we have

$$\begin{aligned} G_\mu x_\mu - i'H_\mu\xi_\mu &= \tfrac{1}{2}P\{(E_\mu+i'F_\mu)(x_\mu+x_\mu')-i'(F_\mu+i'E_\mu)(x_\mu-x_\mu')\}/(1+i') \\ &= P(E_\mu x_\mu + i'F_\mu x_\mu')/(1+i'). \end{aligned}$$

Or writing

$$\mathbf{x} = \sum_1^4 E_\mu x_\mu, \quad \mathbf{X} = \sum_1^4 G_\mu x_\mu, \text{ etc.},$$

$$\mathbf{X} - i'\boldsymbol{\xi} = \frac{P}{1+i'}(\mathbf{x}+i'\mathbf{x}'). \qquad (10\cdot871)$$

Similarly, if

$$\nabla_x = \sum_1^4 E_\mu \partial/\partial x_\mu, \quad \nabla_X = \sum_1^4 G_\mu \partial/\partial x_\mu, \text{ etc.},$$

$$\nabla_X - i'\nabla_\xi = \frac{2P}{1+i'}(\nabla_x + i'\nabla_{x'}). \qquad (10\cdot872)$$

For example, if we have two non-interacting particles of proper mass m which satisfy the wave equation (8·262) without electromagnetic field, viz.

$$(\nabla_x - im)\psi = 0, \quad (\nabla_{x'} - im)\phi = 0,$$

the product wave function Ψ satisfies both equations, and therefore satisfies

$$\{(\nabla_x - im) + i'(\nabla_{x'} - im)\}\Psi = 0.$$

Hence, multiplying by $2P/(1+i')$,

$$(\nabla_X - i'\nabla_\xi - 2iPm)\Psi = 0. \tag{10·88}$$

A noteworthy point is that the vectors $\boldsymbol{x}$, $\boldsymbol{x}'$, and also the vectors $\mathbf{x}$, $\boldsymbol{\xi}$, have to be combined with a quarter period difference of phase represented by i'. There is evidently no other way of obtaining a transformation of the kind required. If we try to combine them without phase difference, we have

$$E_\mu x_\mu + F_\mu x_\mu' = (E_\mu + F_\mu)x_\mu + (E_\mu - F_\mu)\xi_\mu.$$

The trouble then arises that $E_\mu + F_\mu$, $E_\mu - F_\mu$ are singular symbols with eigenvalues 0 and $\pm 2i$. The product of two vectors $(E_1 + F_1)x_1$, $(E_1 - F_1)\xi_1$ in the same direction in the absolute and relative spaces is identically zero. This means that we have made a singular transformation, so that either the external (absolute) configurations or the internal (relative) configurations have shrunk to a point in our representation.

In the transformation here found x_μ and ξ_μ are referred to symbolic frames which are equivalent to one another and to the original frames of x and x'. Thus the relative coordinates can be treated in the same way as the absolute coordinates hitherto treated. It would be more logical to say that the absolute coordinates can be treated in the same way as relative coordinates, for all our experience is concerned with relative positions. But to render the representations similar we have had to measure the coordinates in various complex units. This is seen by writing (10·871) in the form

$$\frac{\mathbf{x}}{1} + \frac{\boldsymbol{\xi}}{e^{\frac{1}{2}i\pi}} = \frac{P}{\sqrt{2}}\left(\frac{\boldsymbol{x}}{e^{\frac{1}{2}i\pi}} + \frac{\boldsymbol{x}'}{e^{-\frac{1}{2}i\pi}}\right). \tag{10·89}$$

There does not seem to be any way of extending the foregoing transformation to particles of unequal mass. On reflection we see that no such extension could be expected. The integrated mass of a particle distributed uniformly all over the universe is a highly artificial conception, and has no close relation to the concentrated masses of classical particles. Particles with different masses (protons and electrons) do not appear in the theory until a later stage of development than that which we have now reached, and their wave functions are not the primitive relativistic wave functions, introduced by Dirac's equation, which we are at present studying.†

In the next section we treat a more familiar transformation of a somewhat different character which concerns particles of unequal mass.

† See § 11·6.

10·9. Relative Coordinates—Unequal Masses.

For two particles of proper mass m, m', the coordinates x_μ of the centroid and the relative coordinates ξ_μ are given by

$$x_\mu = \frac{mx_\mu + m'x_\mu'}{m+m'}, \quad \xi_\mu = x_\mu - x_\mu'. \tag{10·911}$$

Hence
$$\frac{\partial}{\partial x_\mu} = \frac{m}{m+m'}\frac{\partial}{\partial x_\mu} + \frac{\partial}{\partial \xi_\mu}, \quad \frac{\partial}{\partial x_\mu'} = \frac{m'}{m+m'}\frac{\partial}{\partial x_\mu} - \frac{\partial}{\partial \xi_\mu}. \tag{10·912}$$

Writing
$$M = m+m', \quad \mu = mm'/(m+m'), \tag{10·92}$$

we obtain from (10·912) the well-known formula

$$\frac{1}{m}\frac{\partial^2}{\partial x_\mu^2} + \frac{1}{m'}\frac{\partial^2}{\partial x_\mu'^2} = \frac{1}{M}\frac{\partial^2}{\partial x_\mu^2} + \frac{1}{\mu}\frac{\partial^2}{\partial \xi_\mu^2}. \tag{10·93}$$

Introducing momenta, $p_\mu = -i\partial/\partial x_\mu$, $\varpi_\mu = -i\partial/\partial \xi_\mu$, etc., this becomes

$$p_\mu^2/m + p_\mu'^2/m' = P_\mu^2/M + \varpi_\mu^2/\mu. \tag{10·94}$$

In the transformation (10·911), the jacobian

$$\frac{\partial(x_\mu, \xi_\mu)}{\partial(x_\mu, x_\mu')} = 1 \tag{10·95}$$

so that the transformation $x_\mu, x_\mu' \to x_\mu, \xi_\mu$ does not alter volume elements. Normally (10·911) applies to space-coordinates only, the time being the same for the two particles. We can, if occasion arises, extend it formally to four coordinates; but it must then be remembered that the relative time $\tau = t - t'$ is not the time which is ordinarily associated with relative coordinates ξ_μ.

We notice especially that

$$mm' = M\mu = M, \text{ say.} \tag{10·96}$$

We have seen (§ 10·1) that a system is to be regarded as the product, rather than as the sum, of its parts; and from this point of view the product mass M is its leading dynamical characteristic. The two ways of factorising M lead to two modes of dividing the system—either into two ordinary particles m, m', or into an *external particle* M and an *internal particle* μ. These two modes of factorising M are familiar in classical mechanics, especially in celestial mechanics, where μ is called the *reduced mass* associated with the relative orbit.

If $p^2 = p_1^2 + p_2^2 + p_3^2$, etc., we obtain by summing (10·94)

$$p^2/m + p'^2/m' = P^2/M + \varpi^2/\mu. \tag{10·97}$$

For simple plane waves the energy was found in (5·251) to be

$$p_0 = -i\partial/\partial t = (m^2 + p^2)^{\frac{1}{2}} \doteq m + \tfrac{1}{2}p^2/m. \tag{10·981}$$

The conditions described by plane progressive waves are, however, highly abstract, and (10·981) has no immediate interpretation in terms of obser-

vation.† All observed motion is relative motion, and the energies with which we are concerned observationally correspond to relative coordinates ξ_μ. Clearly energy must be defined in such a way that either mode of division of the system gives the same total energy. By (10·97) the function of p, which is conserved in the transformation from P, ϖ to p, p', is p^2/m; the variable part of the energy must be equal or proportional to this. We therefore set the energies equal to

$$\left.\begin{aligned} E = m + \alpha p^2/m, \quad & E' = m' + \alpha p'^2/m', \\ \bar{E} = M + \alpha P^2/M, \quad & \epsilon = \alpha \varpi^2/\mu, \end{aligned}\right\} \tag{10·982}$$

where α is a constant coefficient. These satisfy

$$E + E' = \bar{E} + \epsilon. \tag{10·983}$$

We notice in (10·982) that the relative energy ϵ does not include any rest-energy.

It has commonly been supposed that the energy $m + \alpha p^2/m$ is merely an approximation to $(m^2 + p^2)^{\frac{1}{2}}$, the value of α being $\frac{1}{2}$. That is not so; $m + \alpha p^2/m$ is the exact expression in the conditions with which we are here concerned, and (10·981) is irrelevant. According to the common assumption the variable part of the energy is $\{(m^2 + p^2)^{\frac{1}{2}} - m\}$, and it is taken for granted that the relative energy is also of this form, viz. $\{(\mu^2 + \varpi^2)^{\frac{1}{2}} - \mu\}$. The latter expression is the one which concerns us observationally, since we can only observe relative positions and energies. But

$$\{(m^2 + p^2)^{\frac{1}{2}} - m\} + \{(m'^2 + p'^2)^{\frac{1}{2}} - m'\} \neq \{(M^2 + P^2)^{\frac{1}{2}} - M\} + \{(\mu^2 + \varpi^2)^{\frac{1}{2}} - \mu\}.$$

Thus the sum of the internal and external energies of the system is not equal to the sum of the energies of the constituent particles; and it is clear that the form used is inadmissible.

We shall have more to say on this point in Chapter XIII. Meanwhile the following explanation must suffice. The expression for the energy operator in terms of the momentum operators p_1, p_2, p_3 (the three-dimensional hamiltonian) is a leading dynamical characteristic of a system. What we here find is that the two parts into which we divide a system have not the same dynamical characteristic as the set of plane waves studied in § 5·2. It was scarcely likely that they would; for our first investigations were concerned with abstract conditions, and it is only with the introduction of double wave functions that we begin to make some approach to actuality. In order that the wave equation $H\psi = 0$ of a simple wave vector may be invariant for rotations of space-time, H must be a space vector. Correspondingly, if the wave equation of a double wave vector is $H_2 \Psi = 0$, H_2

† Its application to the observed change of mass with velocity of an electron cannot be described as *immediate*, since the mass is distributed over an infinite wave front. To connect this with the change of mass of an approximately classical particle it is necessary to investigate the transformation of the normalising conditions.

must be a space tensor of the second rank. If we divide H_2 into two parts $H_2' + H_2''$ referring to the two particles, H_2' and H_2'' may or may not be tensors separately; but at any rate they will not be space vectors. Thus the vector form of H, which yields $p_0 = (m^2 + p^2)^{\frac{1}{2}}$, is not applicable to the constituents of a double system; and p^2/m occurs in (10·982) in its own right as a component of a tensor of the second rank, not as an approximation to a component of a vector.

PART II

PHYSICAL APPLICATIONS

CHAPTER XI

THE RIEMANN-CHRISTOFFEL TENSOR

11·1. The Comparison Fluid.

When the position of a body is determined by observation it is referred to landmarks furnished by material objects; but in theoretical formulae the position is supposed to be specified relatively to a geometrical frame. In macroscopic physics the geometrical frame is an admissible substitute for material reference marks; and there is no inconsistency in assuming that the position, momentum, etc., defined by reference to points of a geometrical frame, satisfy the same equations as the position, momentum, etc., measured from material landmarks. But it is not so in wave mechanics. A material frame of reference cannot be equivalent to a geometrical frame. For the material system is specified by a wave function which describes a *probability distribution* of its observable landmarks relative to a geometrical frame. To identify the position and motion of the landmarks with the position and motion of a geometrical frame contradicts the uncertainty principle.

Thus in so far as the equations of wave mechanics relate to anything observable, they contemplate (*a*) a geometrical frame, (*b*) a physical reference object or system of objects, and (*c*) the particular particle or system under discussion. The geometrical frame cannot be omitted without abandoning the whole method of wave mechanics, but it is only an intermediary and not the final reference system for our observations; (*b*) and (*c*) are necessary in order to furnish observable phenomena for measurement. For simplicity the casual reference objects (*b*) used in actual experiments are replaced by an ideal standardised reference object. The fundamental equations of physics necessarily contemplate highly idealised systems in highly idealised conditions; and they select a reference object with simple and symmetrical properties, just as they select for investigation very elementary systems (*c*). But the idealisation must not be carried so far as to substitute something which has not the same relation to our sensory experience. Idealisation is permissible; abstraction is not. The practice in current quantum theory of substituting a sharp geometrical frame for a probability distribution of observable landmarks is clearly indefensible.

Probably it has been thought that the difference between (*a*) and (*b*) would be made insensible by using very massive material landmarks. But a massive system causes curvature of space. Current quantum theory neglects curvature of space, and therefore falls into error either way: either it postulates that a light reference object is used and wrongly neglects its

uncertainty of position and velocity, or it postulates a heavy reference object and wrongly neglects the resulting curvature of space. Perhaps the most important insight, obtained through a combination of relativity theory and wave mechanics, is a realisation that the two alternatives are different forms of the same error.

The magnitude of the error introduced by failing to distinguish (*a*) and (*b*) is a matter for detailed calculation, and will be found in due course. We need only say here that the common impression that it is likely to be trivial is very far from the truth.

The idealised physical reference object, which is implied in current quantum theory, is a fluid permeating all space like an aether. Such an aether is in a sense a materialisation of space; or better, space (the geometrical background or metrical field) should be looked upon as a dematerialised abstraction of the adopted comparison aether. "Dematerialisation" is represented analytically by replacing a probability distribution by a sharp configuration. The uncertainty principle does not allow such replacement until we have removed from the comparison fluid those characteristics which make it accessible to observation—so that it no longer serves the purpose of a reference mark for observational measurements.

We shall call the three constituents of any problem (*a*) the frame, (*b*) the comparison fluid, (*c*) the object system.

In future developments frequent reference will be made to the comparison fluid.† It must be understood that the comparison fluid is not a hypothesis. Still less is it an ascertained feature of the universe. The comparison fluid is a datum of the elementary problems which we treat—like the "frictionless constraints" so often specified as data in problems in elementary mechanics. There is no suggestion that a comparison aether actually exists throughout the universe. But, instead of a comparison aether, irregular distributions of matter exist and furnish landmarks for our observational measurements; the comparison fluid takes the place of these in our elementary equations. From these we can, if necessary, proceed to more complicated equations which take account of the irregularity of the background of our actual experiments.

If anyone is disposed to offer criticism of this view of the equations of wave mechanics, it is necessary that he should state what (in his view) is the physical reference object intended by writers on wave mechanics when they mention positions, momenta, etc. For, if he believes the quantum equations to be true observationally, he must surely be prepared to state what observable system he believes them to be true of. We would remind him that a

† The "standard environment" occasionally referred to in earlier chapters is now replaced by the comparison fluid.

geometrical frame is not observable; nor is it a simplified representation of anything observable; its sharpness is, according to the principles of wave mechanics, incompatible with observability.

The axiom of relativity is that we can only observe relations between physical entities. A relation implies two systems or two parts of a system; it can therefore occur only in connection with a double wave function, the duplicity corresponding to the two ends of the relation.

It is therefore evident that the union of wave mechanics with relativity theory must be based on the theory of double wave tensors. As we pointed out in § 10·75, the recognition that the simple wave tensor is an abstraction, and that the double wave tensor is the primary physical concept, is the counterpart in relativistic wave mechanics of the recognition in macroscopic relativity theory that the interval is the primary physical concept.

This conclusion is emphasised by the fact that the fundamental tensors in relativity theory are symmetrical tensors of the second rank, namely, the metrical tensor $g_{\mu\nu}$ and the energy tensor $T_{\mu\nu}$. In symbolic notation these yield double wave tensors $\Sigma g_{\mu\nu}E_\mu F_\nu$ and $\Sigma T_{\mu\nu}E_\mu F_\nu$. On account of its antisymmetrical properties the Riemann-Christoffel tensor (although it is of the fourth rank) can also be expressed as a double wave tensor $\Sigma B_{\mu\nu\sigma\epsilon}E_{\mu\epsilon}F_{\nu\sigma}$.

In the more elementary problems, the double wave tensors refer to the object system and comparison fluid, and specify the probability distribution of their combined configurations. What are ordinarily described as properties of the object system must be understood to mean properties of the combination object system + comparison fluid.

11·2. Linked Rotations.

In general relativity theory, a vector A_μ carried by parallel displacement round the perimeter of a surface element $dS^{\nu\sigma}$ receives an increment†

$$dA_\mu = \Sigma_\nu \Sigma_\sigma \Sigma_\epsilon \tfrac{1}{2} B_{\mu\epsilon\nu\sigma} A^\epsilon dS^{\nu\sigma}, \qquad (11\cdot21)$$

where B is the Riemann-Christoffel tensor. We have changed the usual order of writing the suffixes (which would have been $B_{\mu\nu\sigma\epsilon}$), since it is found to be very inconvenient in the present theory. We shall adopt local Galilean coordinates (natural coordinates); since real time is used, it is necessary to pay attention to the upper and lower positions of the suffixes.

Let $dS^{\nu\sigma}$ be a circle of infinitesimal radius r in the plane $x_\nu x_\sigma$, described in the direction $x_\nu \to x_\sigma$. A vector A_ϵ along the x_ϵ axis, by receiving an increment dA_μ along the x_μ axis, is rotated in the $x_\mu x_\epsilon$ plane through an angle $d\theta_{\mu\epsilon}{}'$ given by

$$-dA_\mu = A^\epsilon d\theta_{\mu\epsilon}{}',$$

so that by (11·21)

$$d\theta_{\mu\epsilon}{}' = -\pi r^2 B_{\mu\epsilon\nu\sigma}. \qquad (11\cdot22)$$

† *Mathematical Theory of Relativity*, equation (84·1).

The factor $\frac{1}{2}$ has disappeared because the surface element has two components $dS^{\nu\sigma} = -dS^{\sigma\nu} = \pi r^2$.

The components of B therefore specify a linkage of two rotations, the one in the $\nu\sigma$ plane and the other in the $\mu\epsilon$ plane, the first being a rotation of the position vector of the point considered and the second being a rotation of an arbitrary vector situated at the point considered. In physical applications we regard the point as occupied by an object system, and the rotations refer respectively to the position vector of the object system and to a vector regarded as intrinsic in the object system. We shall call the two vectors the *position vector* and the *intrinsic vector*.

In (11·22) the angle of rotation of the position vector is $\theta^{\nu\sigma} = 2\pi$. The ratio of the linked rotations is therefore

$$\theta_{\mu\epsilon}{}'/\theta^{\nu\sigma} = -\tfrac{1}{2}r^2 B_{\mu\epsilon\nu\sigma}. \qquad (11{\cdot}231)$$

Or, expressing both angles contravariantly,

$$\theta'^{\mu\epsilon}/\theta^{\nu\sigma} = -\tfrac{1}{2}r^2 B^{\mu\epsilon}{}_{\nu\sigma}. \qquad (11{\cdot}232)$$

In "de Sitter space-time", i.e. space-time spherical in its space dimensions and hyperbolic in the time dimension, the Riemann-Christoffel tensor is

$$B^{\mu\epsilon}{}_{\nu\sigma} = (\delta_\nu{}^\mu \delta_\sigma{}^\epsilon - \delta_\sigma{}^\mu \delta_\nu{}^\epsilon)/R^2. \qquad (11{\cdot}233)$$

This vanishes unless the planes $\mu\epsilon$, $\nu\sigma$ coincide. If $\mu, \epsilon = \nu, \sigma$ the value is $1/R^2$. Hence, in de Sitter space-time the two rotations are in the same plane, and their ratio is

$$\theta'^{\nu\sigma}/\theta^{\nu\sigma} = -\tfrac{1}{2}r^2/R^2, \qquad (11{\cdot}234)$$

r being the length of the position vector (treated as infinitesimal) and R the radius of space-time.

An illuminating point of view is obtained by interpreting θ' as the rotation of a comparison fluid. The intrinsic vector A_μ has been parallelly displaced; this implies that, although its mathematical specification is altered, it has in some physical sense suffered no real change. It is still open to us to define the observational significance of the concept "real change"; this definition will decide the way in which the analytical theory of parallel displacement is to be utilised in the study of physical phenomena. We shall define *real change* to mean *observable change*. The orientation of the vector A_μ in the geometrical frame is not observable; what we observe is its orientation in the comparison fluid which is the idealised substitute for material landmarks. Thus A_μ will suffer no observable change if its direction remains fixed in the comparison fluid. Being fixed in the fluid it undergoes the same rotation θ' as the comparison fluid relative to the (unobservable) frame.

The linkage of a rotation θ' of the comparison fluid to a rotation θ of the position vector of the object system is conceived dynamically as a recoil of the comparison fluid. If the fluid is assigned a moment of inertia $2mR^2$,

(11·234) expresses that when the object system is given an angular momentum $mr^2 d\theta/dt$ the comparison fluid recoils with equal and opposite angular momentum $2mR^2 d\theta'/dt$. In non-uniform space-time the comparison fluid has (as we should expect) a momental ellipsoid with unequal axes, so that the recoil angular velocity $d\theta'/dt$ is not necessarily in the same plane as the recoil angular momentum, and therefore not necessarily in the same plane as $d\theta/dt$. Thus components of $B_{\mu\epsilon\nu\sigma}$ for which $\mu, \epsilon \neq \nu, \sigma$ are introduced. But we need not unduly stress this dynamical model. The R.C. tensor specifies directly the kinematical recoil of the comparison fluid corresponding to every possible cyclic displacement of the object system. If we introduce dynamical conceptions, we must attribute to the comparison fluid whatever dynamical characteristics are necessary to satisfy this specification, whether they are illustrated by familiar dynamical models or not.

It may be noticed that a completely arbitrary connection between $\theta_{\mu\epsilon}'$ and $\theta_{\nu\sigma}$ cannot be represented by an R.C. tensor; the arbitrary connection would have 36 disposable constants, whereas the R.C. tensor has 20 independent components.†

From the standpoint of macroscopic theory the foregoing is a highly speculative interpretation of the R.C. tensor. There is in fact an insuperable objection to adopting it in macroscopic relativity theory, which will be explained in § 11·3. But in microscopic theory we approach it in a different way; the speculative taint is removed; and the aforementioned objection disappears as soon as we substitute displacement by wave propagation for the classical conception of displacement.

In wave mechanics the comparison fluid presents itself as a datum of the problem; without it the mathematical equations could have no relation to observable phenomena. We *observe* displacements of the object system relative to the comparison fluid; but the methods of mathematical physics require that we should use in our equations displacements relative to a geometrical frame. For convenience we call displacements relative to the frame *absolute* (i.e. conceived as absolute). We have therefore to analyse an observed relative displacement into absolute displacements of the object system and comparison fluid, before we can apply the theoretical equations. This partition can only be decided by convention. The convention is arbitrary in the first place; but it may later be limited so as to fulfil conditions which simplify the resulting formulae.

We have therefore to draw up a scheme of partition specifying the two absolute motions which correspond to any given relative motion; or equivalently the scheme will specify the absolute motion of the comparison fluid corresponding to any given absolute motion of the object system. We have

† We confine ourselves to Riemannian geometry. In affine geometry a generalised curvature tensor with 36 independent constants is substituted.

seen that (for rotational displacements) such a scheme of linkage can be expressed by a R.C. tensor $B_{\mu\epsilon\nu\sigma}$.

Defined in this way as a scheme of partition, $B_{\mu\epsilon\nu\sigma}$ has no immediate relevance to curvature of space. But the analysis at the beginning of this section shows that, if we ascribe to the unobservable frame a *metric* whose R.C. tensor is $B_{\mu\epsilon\nu\sigma}$, the *partition* represented by $B_{\mu\epsilon\nu\sigma}$ becomes automatic —on the understanding that displacement without observable change is represented geometrically by parallel displacement. "Observable", of course, means observable under the ideal conditions of the problem, namely that the only physical systems present are the object system and the comparison fluid.

This double aspect of $B_{\mu\epsilon\nu\sigma}$ is the most essential link between relativity theory and wave mechanics. By it energy, momentum and mass, which appear in relativity theory as components of curvature of space-time, become identified with coefficients involved in the partition of relative displacements into absolute displacements, and therefore in the analysis of the double wave functions (which contain observational information) into the abstract simple wave functions which form the ordinary starting point of wave mechanics. In this way energy, momentum and mass get their footing in wave mechanics.

The double interpretation of $B_{\mu\epsilon\nu\sigma}$ also exhibits the way in which the geometrical conception of mass as curvature becomes translated into a dynamical conception. The coefficients of the scheme of partition determine the recoil of the comparison fluid. An object particle whose displacements produce a large recoil is regarded dynamically as having a large mass. We see from (11·231) that if the R.C. tensor, and therefore the mass, is multiplied by any factor the recoil is multiplied by the same factor.

11·3. Displacement by Wave Propagation.

In obtaining (11·231) we assumed that the change dA_μ recognised at the end of a complete cycle had occurred evenly during the cycle. In macroscopic physics it is impossible to admit this. When the object system receives a displacement dx_1, we do not know whether the infinitesimal displacement is going to form part of a circuit in the x_1x_2 plane or the x_1x_3 plane, and it is therefore impossible to assign the plane of the recoil $d\theta'$ of the comparison fluid. General relativity theory has therefore adopted a different representation. It admits a rotation (usually accompanied by strain) of a vector parallelly displaced through dx_1; but the rotation is described by a *non-tensor* quantity—a 3-index symbol. The change assigned is thus dependent on the coordinate system employed; it cannot be pictured absolutely, or attributed to an objective rotation of the comparison fluid.

This difficulty does not arise in wave mechanics, because we do not con-

template displacements of the classical type. The nearest approximation to classical displacement is propagation of a wave packet. But the wave packet is analysed into elementary waves, each of which fills the whole of space. An elementary displacement corresponds to the propagation of one of these waves. To each elementary wave there is a corresponding recoil of the comparison fluid, and the recoil due to the propagation of the wave packet is the resultant of these.

Wave analysis removes all ambiguity as to the plane of rotation, because it is a wave front, not a point, which is displaced. The free motion of a particle or wave packet is commonly analysed into "infinite plane waves"; but these must be modified to fit finite space. In spherical space-time we can mark the wave front on the sphere, and obtain the wave propagation by rotating the sphere in the plane defined by the wave normal and the radius of curvature of space-time. For particles bound in an atom the analysis into elementary states gives waves representing angular motion; these correspond to rotations in three-dimensional space.

It would be difficult to extend to all kinds of irregular space the conception of wave propagation as a simple rotation. In general the normal to space-time is six-dimensional. But wave mechanics evades all such complications by analysing physical systems into steady states in which the motion is characterised by constant angular momentum. Changes of distribution which cannot be represented by rotation (in the most generalised sense, i.e. displacement without intrinsic change in any of the planes of the symbolic frame) are treated by perturbation methods, and are not recognised as spatio-temporal displacements. If wave mechanics is unable to describe the change from configuration A to configuration B in terms of rotation, it represents the change as spatially discontinuous; that is to say, it examines the transfer of probability from A to B without representing the system as having passed through configurations intermediate between A and B. Such changes are called *transitions*, not displacements.

The result is that every spatio-temporal displacement contemplated in wave mechanics is a rotation, or is analysed into rotations. Indeed the theory developed in the previous chapters makes no provision for displacements which are not the manifestation of rotation; and every continuous transformation or change of configuration is associated with a symbol defining a plane in a symbolic frame. Since we always know the plane of rotation associated with a displacement, the difficulty encountered in macroscopic physics does not arise.

The obvious example in microscopic physics of an object system describing a small circle $dS^{\nu\sigma}$ is furnished by the electron in a hydrogen atom. We can contemplate the displacement of the electron during a time short compared with the time of revolution; but the displacement can only be described by

an angle $d\theta_{\nu\sigma}$, not by a line element dx_ν, because the position angle of the electron at any instant is indeterminate.

If then we consider an electron with constant angular momentum in the plane $\nu\sigma$,

$$mr^2\frac{d\theta}{dt}=\frac{nh}{2\pi}, \tag{11·31}$$

the recoil of the comparison fluid in the plane $\nu\sigma$ will be, by (11·232),

$$-\frac{d\theta'}{dt}=\tfrac{1}{2}r^2B^{\nu\sigma}{}_{\nu\sigma}\frac{d\theta}{dt}=\frac{nh}{4\pi m}B^{\nu\sigma}{}_{\nu\sigma}, \tag{11·32}$$

or, assuming spherical space,

$$-\frac{d\theta'}{dt}=\frac{nh}{4\pi mR^2}. \tag{11·33}$$

It is desirable to meet at this stage the criticism that an effect of the order (11·33) is utterly trivial. The recoil $d\theta'/dt$ applies to a comparison fluid filling the whole universe. There are, say, N other electrons in the universe, whose contributions to the total recoil must be added together. If (as seems to be generally implied in the elementary formulae) these are all supposed to be doing the same thing, the total recoil is

$$-\frac{d\theta'}{dt}=\frac{nh}{4\pi mR'^2}, \tag{11·34}$$

where

$$R'=R/\sqrt{N}\sim 3\,.\,10^{-13}\,\text{cm}. \tag{11·35}$$

according to the value of N found in Chapter XIV. This is by no means trivial; and outside the region of the nucleus ($r>10^{-12}$ cm.), $d\theta'/dt$ is greater than $d\theta/dt$.

If, on the other hand, we suppose that (11·31) refers to one electron only, it is right that the recoil (11·33) should be of the same order of observational insignificance as the motion which occasions it. To detect that one of the N electrons in the universe has acquired a rotation $d\theta/dt$ can scarcely be said to be within reach of practical observation. It is true that we *might* observe a reasonably large effect, but the chances are N to 1 that we shall observe nothing; for there is nothing in the mathematical formulation to indicate that the electron referred to in (11·31) is the one which the observer is watching.

In natural coordinates the continuum of space and imaginary time is isotropic. We have found that, for a spatial displacement of the object system, the comparison fluid recoils in the opposite direction; and the same holds for an imaginary time displacement. If we substitute real time, the sign of the corresponding components of the R.C. tensor is reversed; and the real time displacement of the comparison fluid is in the same direction as that of the object system. In our ordinary outlook, objects travel forward in time together. More particularly in a measurement involving an object

system and physical reference system, it is implied that both systems are observed simultaneously in the time reckoning adopted. We have seen that the scheme of linkage of displacements is initially an arbitrary convention, and for the present we shall treat it as arbitrary; but at the appropriate stage we shall specialise it so that the time displacements (in the time reckoning adopted) of the object system and comparison fluid are equal, as the ordinary outlook assumes.

11·4. The Riemann-Christoffel Matrix.

We shall now express the R.C. tensor as a wave tensor. In order to utilise the relation between wave tensors and space tensors found in our previous work, we must take the fourth coordinate to be imaginary time. Accordingly we use local rectangular coordinates with $g_{\mu\nu} = -\delta_\mu{}^\nu$; and there is no longer any distinction between $B^{\mu\epsilon}{}_{\nu\sigma}$ and $B_{\mu\epsilon\nu\sigma}$.

We define the *Riemann-Christoffel Matrix* to be†

$$(B)_{\alpha\gamma\beta\delta} = \Sigma B_{\mu\epsilon\nu\sigma} (E_{\mu\epsilon})_{\alpha\beta} (E_{\nu\sigma})_{\gamma\delta}. \qquad (11{\cdot}41)$$

Since the suffixes of the E-symbols run from 0 to 5, we require components of $B_{\mu\epsilon\nu\sigma}$ additional to those which make up the ordinary R.C. tensor. The physical interpretation of the extra components will be found in due course; but their primary significance is that they specify linkages of the rotations of the object system and comparison fluid in the corresponding planes as in § 11·2.

Consider a factorisable matrix (as in (10·413))

$$(B)_{\alpha\gamma\beta\delta} = U_{\alpha\gamma} V_{\beta\delta}. \qquad (11{\cdot}421)$$

Treating U and V as double vectors in the CD frame, we have by (10·24)

$$B_{\mu\epsilon\nu\sigma} = \tfrac{1}{16} V_{\alpha\gamma} (E_{\mu\epsilon})_{\alpha\beta} (E_{\nu\sigma})_{\gamma\delta} U_{\beta\delta}. \qquad (11{\cdot}422)$$

By the symmetry property of the R.C. tensor, $B_{\mu\epsilon\nu\sigma} = B_{\nu\sigma\mu\epsilon}$. It follows from (11·41) that $(B)_{\alpha\gamma\beta\delta} = (B)_{\gamma\alpha\delta\beta}$, so that, by (11·421),

$$U_{\alpha\gamma} V_{\beta\delta} = U_{\gamma\alpha} V_{\delta\beta}$$

or $UV = \overline{U}\,\overline{V}$. This is satisfied if U and V are both symmetrical or both antisymmetrical. We shall assume for the present that both are symmetrical; the antisymmetrical case is treated in § 11·5.

There is a simple generalisation of the identity (5·41), which extends it to any two four-valued quantities ψ, χ, namely

$$\sum_{\sigma=0}^{\sigma=5} \{(E_{\mu\sigma}\psi)_\alpha (E_{\mu\sigma}\chi)_\gamma + (E_{\mu\sigma}\psi)_\gamma (E_{\mu\sigma}\chi)_\alpha\} - \{(E_{16}\psi)_\alpha (E_{16}\chi)_\gamma + (E_{16}\psi)_\gamma (E_{16}\chi)_\alpha\} \equiv 0. \qquad (11{\cdot}43)$$

† It will be seen that the suffixes are so arranged that the R.C. tensor is represented in a CD frame (§ 10·4). This has a slight advantage in certain later developments.

This is proved in the same way as (5·41),† and holds for the six pentads given by $\mu = 0, 1, 2, 3, 4, 5$. Let

$$U_{\alpha\gamma} = \psi_\alpha \chi_\gamma + \chi_\alpha \psi_\gamma. \tag{11·441}$$

Then (11·43) becomes

$$\{\Sigma_\sigma (E_{\mu\sigma})_{\alpha\beta} (E_{\mu\sigma})_{\gamma\delta} - (E_{16})_{\alpha\beta} (E_{16})_{\gamma\delta}\} U_{\beta\delta} = 0. \tag{11·442}$$

Multiplying initially by $V_{\alpha\gamma}$ and using (11·422), we obtain

$$\Sigma_\sigma B_{\mu\sigma\mu\sigma} - B_{1616} = 0. \tag{11·451}$$

Also, multiplying initially by $V_{\epsilon\gamma} (E_{\mu\nu})_{\epsilon\alpha}$, we obtain

$$\Sigma_\sigma B_{\nu\sigma\mu\sigma} = 0 \quad (\mu \neq \nu). \tag{11·452}$$

This last reduction depends on the rule $(E_{\mu\nu})_{\epsilon\alpha} (E_{\mu\sigma})_{\alpha\beta} = (E_{\nu\sigma})_{\epsilon\beta}$ when $\sigma \neq \nu$. The term for which $\sigma = \nu$ reduces to $V_{\epsilon\gamma} (-1)_{\epsilon\beta} (E_{\mu\nu})_{\gamma\delta} U_{\beta\delta}$, which cancels the last term $-V_{\epsilon\gamma} (iE_{\mu\nu})_{\epsilon\beta} (i)_{\gamma\delta} U_{\beta\delta}$, owing to the symmetrical property of U and V.

We now express the results (11·451), (11·452) in terms of the Einstein tensor $G_{\mu\nu}$ and invariant G. Since $g_{\mu\nu} = -\delta_\mu{}^\nu$,

$$\left.\begin{aligned} -G_{\mu\nu} &= B_{\mu1\nu1} + B_{\mu2\nu2} + B_{\mu3\nu3} + B_{\mu4\nu4}, \\ -G &= G_{11} + G_{22} + G_{33} + G_{44}. \end{aligned}\right\} \tag{11·461}$$

We use the above definition of $G_{\mu\nu}$ to define new components G_{55} and G_{00}. Then (11·451) becomes

$$-G_{\mu\mu} + B_{\mu5\mu5} + B_{\mu0\mu0} = B_{1616} \tag{11·462}$$

or, by the antisymmetrical properties of $B_{\mu\epsilon\nu\sigma}$,

$$-G_{\mu\mu} + B_{5\mu5\mu} + B_{0\mu0\mu} = B_{1616}. \tag{11·463}$$

Summing (11·463) for $\mu = 1, 2, 3, 4$, we obtain

$$G - G_{55} - G_{00} = 4B_{1616}. \tag{11·464}$$

But by (11·462)

$$-G_{55} + B_{5050} = B_{1616}, \quad -G_{00} + B_{0505} = B_{1616}. \tag{11·465}$$

Hence, writing

$$B_{5050} = B_{0505} = \lambda \tag{11·466}$$

we obtain from (11·464) and (11·465)

$$B_{1616} = \tfrac{1}{2}(G - 2\lambda). \tag{11·47}$$

Thus (11·451) and (11·452) become

$$\Sigma_\sigma B_{\mu\sigma\nu\sigma} = -\tfrac{1}{2} g_{\mu\nu} (G - 2\lambda).$$

Whence, by (11·461),

$$\begin{aligned} B_{\mu5\nu5} + B_{\mu0\nu0} &= G_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} (G - 2\lambda) \\ &= -8\pi T_{\mu\nu}, \end{aligned} \tag{11·48}$$

by the usual relativity formula for the energy tensor $T_{\mu\nu}$.‡

† The reader who wishes to check the calculation may find it helpful to refer to a later result, (11·54), (11·551).

‡ *Mathematical Theory of Relativity*, equation (54·71). The constant λ is at our disposal, and its value must be fixed by a convention (see § 13·1). Here the value of λ is chosen so that the macroscopic energy tensor may have a simple connection with the energy, etc. defined in wave mechanics; it will, in fact, appear as we proceed, that the mechanical identifications adopted in current wave mechanics presuppose that energy, etc. are

If the axes 5 and 0 are treated as invariant, (11·48) is a tensor equation, and therefore holds for any orientation of the axes in four dimensions.

We have proved (11·48) for a factorisable R.C. matrix, but owing to the linearity of the equations the proof can be immediately extended to non-factorisable matrices, which must be expressed as the sum of a number of products UV in (11·421). As already stated, it is postulated that U and V are symmetrical matrices; when they are antisymmetrical some additional terms appear which will be investigated in § 11·5.

By (11·48) the energy tensor is composed of two parts, which we shall distinguish as a kinematical part $(T_{\mu\nu})_k$ and an electrical part $(T_{\mu\nu})_e$, namely,

$$8\pi (T_{\mu\nu})_k = -B_{\mu 5\nu 5}, \quad 8\pi (T_{\mu\nu})_e = -B_{\mu 0\nu 0}. \tag{11·49}$$

Considering the kinematical part, the component $T_{\mu\nu}$ determines the rotation of the comparison fluid in the $\mu 5$ plane linked to a rotation of the position vector of the object system in the $\nu 5$ plane. These are the rotations which correspond to translations in space-time (§ 6·2). Accordingly:

The kinematical energy tensor consists of those components of the whole Riemann-Christoffel tensor which specify linkages of translations *in four dimensions.*

It is this part of the tensor which is used in dissecting a relative motion of translation into separate absolute motions of the object system and comparison fluid.

We now see why this part of the R.C. tensor is sufficient for macroscopic mechanics. The "continuous matter" treated in macroscopic relativity theory is supposed to be such that its kinematics can be described by macroscopically continuous displacements in four dimensions. The theory makes no provision for microscopic vorticity. Atomic physics, on the other hand, is intimately concerned with vorticity or angular momentum on a microscopic scale. Relativity theory can take account of microscopic motions (e.g. heat energy of a gas), but only as signless quantities—mean-square values. If the atoms were spinning preponderantly in one direction, there would be no indication of this in the motions as averaged for macroscopic purposes.

Thus the mechanics of the atom cannot be treated on the basis of the energy tensor $T_{\mu\nu}$ alone. The complete R.C. tensor $B_{\mu\epsilon\nu\sigma}$, including components with suffixes 0 and 5, plays the part of a general mechanical tensor applicable to atomic as well as macroscopic problems. The macroscopic energy tensor forms part of it, $B_{\mu 5\nu 5}$. In atomic problems it is generally

measured from the zero defined by the values of λ given here and in § 11·5. In cosmical problems the value is fixed by another independent convention which has secured general recognition; this cosmical value will not necessarily agree with the value here used, as we shall find in (11·592) and (11·593).

more convenient to handle the complete tensor in its wave tensor form, i.e. as the R.C. matrix $(B)_{\alpha\gamma\beta\delta}$.

It has always been difficult to see the physical significance of the process of contraction used in obtaining the energy tensor from the R.C. tensor. We have no physical insight into what we are doing when we add together $B_{1212}+B_{1313}+B_{1414}$ to form $-G_{11}$. We now see that from the physical point of view the energy tensor T_{11} is not specially related to this sum, but is another component B_{1515} of the same tensor. The summation is part (but only part) of a process of calculating T_{11} when other components are known—depending on an identity satisfied by the components. But equally B_{1212} could be calculated from the same identity if the other components were known.

In the generalised field theory† I have derived the fundamental tensor ${}^*B_{\mu\nu\sigma}{}^{\epsilon}$ from the theory of relation structure. Since only a portion of the tensor (viz. the part preserved in the contracted tensor ${}^*G_{\mu\nu}$) was actually used in specifying the electromagnetic-gravitational field, we seemed to have "dragged up from below a certain amount of apparently useless lumber" (*loc. cit.* § 99). It was natural to hope that the unused part might ultimately be needed in the representation of microscopic structure; though at the time no progress could be made in this direction. We now see that this hope is fulfilled, and the full tensor ${}^*B_{\mu\nu\sigma}{}^{\epsilon}$ is utilised in microscopic theory. (${}^*B_{\mu\nu\sigma}{}^{\epsilon}$ differs from $B_{\mu\nu\sigma}{}^{\epsilon}$ in providing for a macroscopic electromagnetic field, which would be an unnecessary complication at this stage of our investigation.) There is accordingly thorough continuity between the present theory and the field theory developed in my earlier book.

Field theories which are based on "five-dimensional relativity", i.e. theories which introduce a curvature tensor in five dimensions, have not so simple a connection with the present theory. $B_{\mu\epsilon\nu\sigma}$ has the double aspect of a partition tensor and a curvature tensor; as a partition tensor it includes components corresponding to two additional suffixes 5, 0; but the identification with a curvature tensor applies only to the suffixes 1, 2, 3, 4, and in the present theory we do not extend this identification. For such an extension it would be necessary to introduce a fictitious extension of the distribution over a fifth or sixth dimension, so as to provide corresponding Christoffel brackets and components of curvature. Even if this were formally possible, it would be a retrograde development, undoing the advance made in §§ 6·3, 6·4. Whether the number of suffixes is four or six, $B_{\mu\epsilon\nu\sigma}$ is a function of four coordinates only. It is to be remembered that wave mechanics is a statistical theory. If we introduce fictitious dimensions, the generalised formulae will determine steady states in five or six dimensions; but these are not the states with which we are concerned in physics.

† *Mathematical Theory of Relativity*, Chapter VII, Pt. II.

As an elementary example, let the object system be a particle with an exact momentum. Let us choose the time direction so that it is at rest. Its energy tensor then reduces to a single component $T_{44}=\rho_0$, where ρ_0 is the proper density of its probability distribution. By (11·231) a time displacement $\delta\theta_{45}$ of the object particle corresponds to a time displacement $\delta\theta_{45}'$ of the comparison fluid, given by

$$\begin{aligned}\delta\theta_{45}' &= -\tfrac{1}{2}r^2B_{4545}\delta\theta_{45}=4\pi r^2T_{44}\delta\theta_{45} \quad \text{by (11·49)}\\ &=4\pi r^2\rho_0.\delta\theta_{45}. \qquad (11\cdot495)\end{aligned}$$

For any other direction of displacement $\delta\theta'=0$.

It may seem surprising that in this case the displacement of the particle in spatial directions causes no recoil of the comparison fluid. But it must be remembered that the conditions are highly idealised. By taking the momentum vector to be exact the position of the particle becomes entirely unobservable. Therefore the problem of analysing an observed spatial displacement into absolute displacements of the object particle and comparison fluid does not arise. To represent observed position in space we must introduce a wave packet; the momentum then ceases to be wholly in one direction, and we cannot choose the time direction so that $T_{\mu\nu}$ reduces to a single component. There will accordingly be components T_{11}, etc., specifying the way in which the observed spatial displacement is to be partitioned.

Returning to a particle with an exact momentum vector, the particle is distributed with even probability over the wave front of its waves. Displacement in the wave front is merely a transfer of our attention from one point to another, which has no dynamical reaction on the comparison fluid. Displacement normal to the wave front (relative to the frame) is a physical change of the conditions, involving a linked displacement of the comparison fluid. We see from (11·495) that, when the object particle moves forward in time, the comparison fluid moves in the same direction. We have pointed out that the usual outlook, which has become incorporated in the equations of physics, requires that an object system and its physical reference objects should always move together in time—that the reference system should be a simultaneous one.† We have therefore the condition

$$R\delta\theta_{45}'=r\delta\theta_{45}, \qquad (11\cdot496)$$

where $R\delta\theta_{45}'$ is the displacement of the comparison fluid in linear measure. By (11·495) and (11·496)

$$\rho_0=1/4\pi Rr. \qquad (11\cdot497)$$

When we said above that a time displacement of the particle was a physical change of the conditions, we regarded the displacement as being applied to

† This refers primarily to coordinate time, and so depends on the choice of coordinates. But, in the present simple example, if the two systems agree in coordinate time they will agree in proper time.

the particle and not to anything else—just as one regards space displacements. But it is a rooted habit of thought that a time displacement applied to one particle automatically applies to the rest of the universe—so that time displacement also is a mere transfer of our attention from the present to a future moment. We here reach the same outlook in a more legitimate way by so choosing the linkage of displacement that the object particle automatically carries the comparison fluid with it in the time direction.

As another elementary example, suppose that the object system is identical with the comparison fluid; so that the object is its own standard of reference. Since the object cannot change relatively to its simultaneous self, it will be described observationally as uniform and unchanging. But a uniform unchanging distribution of matter is an Einstein universe. Setting $\delta\theta_{45}' = \delta\theta_{45}$ in (11·495) to express that the object is always compared with its simultaneous self, we have

$$\rho_0 = 1/4\pi r^2, \tag{11·498}$$

which is a well-known formula for the density in an Einstein universe of radius r.

11·5. The Dual Riemann-Christoffel Tensor.

By § 10·4 the R.C. matrix can be resolved in alternative ways. Let

$$(B)_{\alpha\gamma\beta\delta} = \Sigma B_{\mu\epsilon\nu\sigma}(E_{\mu\epsilon})_{\alpha\beta}(E_{\nu\sigma})_{\gamma\delta} = \Sigma b_{\mu\epsilon\nu\sigma}(E_{\mu\epsilon})_{\alpha\gamma}(E_{\nu\sigma})_{\beta\delta} \tag{11·51}$$

or, in the notation of § 10·4,

$$(B) = \Sigma B_{\mu\epsilon\nu\sigma} C_{\mu\epsilon} D_{\nu\sigma} = \Sigma b_{\mu\epsilon\nu\sigma} E_{\mu\epsilon} F_{\nu\sigma}. \tag{11·52}$$

We call $b_{\mu\epsilon\nu\sigma}$ the *dual R.C. tensor.*

An interesting case is when the dual R.C. tensor consists of a single component $b_{1616} = b$. Then, by (10·494),

$$(B) = bE_{16}F_{16} = \tfrac{1}{4}b\Sigma(\pm C_\sigma D_\sigma), \tag{11·531}$$

the minus sign applying to time-like matrices. Hence, by (11·52),

$$B_{\mu\epsilon\mu\epsilon} = \pm\tfrac{1}{4}b, \tag{11·532}$$

the minus sign applying when $E_{\mu\epsilon}$ is time-like.

Our first impression is that (11·532) is the R.C. tensor for de Sitter space-time (11·233), whose quadratic curvature R^{-2} is positive for space dimensions and negative for time dimensions. But in our present coordinate system x_4 is imaginary time; the ambiguity of sign in (11·532) is eliminated when we substitute real time rotations for imaginary time rotations; so that (11·532) represents a world completely isotropic in space and *real* time.

De Sitter space-time is obtained by taking the dual R.C. tensor to consist of a single component $b_{4545} = b$; so that

$$(B) = bE_{45}F_{45} = \tfrac{1}{4}b\Sigma(\pm C_\sigma D_\sigma), \tag{11·533}$$

the minus sign applying to $\sigma = 01, 02, 03, 04, 05, 16$ by (10·495). The plus sign applies to all the components of the ordinary four-dimensional R.C.

tensor; these accordingly agree with the values for a de Sitter space-time of radius given by $\frac{1}{2}b = R^{-2}$.

By (11·421), (11·51) and (11·533)

$$U_{\alpha\gamma}V_{\beta\delta} = (B)_{\alpha\gamma\beta\delta} = b\,(E_{45})_{\alpha\gamma}(E_{45})_{\beta\delta}.$$

Thus U and V in this case consist of single components $E_{45}u_{45}$, $E_{45}v_{45}$, $(u_{45}v_{45} = b)$. Since E_{45} is an antisymmetrical matrix, U and V are antisymmetrical.

Before proceeding further with the investigation of de Sitter space-time, it is necessary to extend the theory given in § 11·4 which treats only the symmetrical case. Let

$$Q_{\alpha\gamma} = \Sigma_\sigma\,(E_{0\sigma}\psi)_\alpha(E_{0\sigma}\chi)_\gamma - (E_{16}\psi)_\alpha(E_{16}\chi)_\gamma. \tag{11·54}$$

According to (11·43) $Q_{\alpha\gamma} = -Q_{\gamma\alpha}$. Using the standard matrices (3·27) for the pentad $E_{0\sigma}$, we find by direct calculation that $\frac{1}{2}Q_{\alpha\gamma}$ is the matrix

$$\begin{array}{c|cccc} \gamma \downarrow \backslash \alpha \rightarrow & & & & \\ & 0 & \psi_2\chi_1 - \psi_1\chi_2 & \psi_2\chi_4 - \psi_4\chi_2 & \psi_4\chi_1 - \psi_1\chi_4 \\ & \psi_1\chi_2 - \psi_2\chi_1 & 0 & \psi_3\chi_2 - \psi_2\chi_3 & \psi_1\chi_3 - \psi_3\chi_1 \\ & \psi_4\chi_2 - \psi_2\chi_4 & \psi_2\chi_3 - \psi_3\chi_2 & 0 & \psi_4\chi_3 - \psi_3\chi_4 \\ & \psi_1\chi_4 - \psi_4\chi_1 & \psi_3\chi_1 - \psi_1\chi_3 & \psi_3\chi_4 - \psi_4\chi_3 & 0 \end{array}$$

This is equivalent to

$$\begin{aligned} \tfrac{1}{2}Q_{\alpha\gamma} &= \psi_\alpha\chi_\gamma - \psi_\gamma\chi_\alpha + (-\psi_3\chi_1 - \psi_4\chi_2 + \psi_1\chi_3 + \psi_2\chi_4)(E_{45})_{\alpha\gamma} \\ &= U_{\alpha\gamma} + \chi^* E_{45}\psi\,.(E_{45})_{\alpha\gamma}, \end{aligned}$$

where $U_{\alpha\gamma} = \psi_\alpha\chi_\gamma - \psi_\gamma\chi_\alpha$. Since

$$\chi_\epsilon(E_{45})_{\epsilon\zeta}\psi_\zeta = -\psi_\epsilon(E_{45})_{\epsilon\zeta}\chi_\zeta = -\tfrac{1}{2}(E_{45})_{\epsilon\zeta}U_{\epsilon\zeta}$$

the result is

$$\tfrac{1}{2}Q_{\alpha\gamma} = U_{\alpha\gamma} - \tfrac{1}{2}(E_{45})_{\epsilon\zeta}U_{\epsilon\zeta}\,.(E_{45})_{\alpha\gamma}. \tag{11·551}$$

If $V_{\beta\delta}$ is another antisymmetrical matrix, we obtain a corresponding R.C. matrix

$$(B)_{\alpha\gamma\beta\delta} = U_{\alpha\gamma}V_{\beta\delta} = \tfrac{1}{2}Q_{\alpha\gamma}V_{\beta\delta} + \tfrac{1}{2}(E_{45})_{\epsilon\zeta}U_{\epsilon\zeta}\,.(E_{45})_{\alpha\gamma}V_{\beta\delta}. \tag{11·552}$$

Since U and V are both antisymmetrical, $(B)_{\alpha\gamma\beta\delta}$ is antisymmetrical in α and γ and in β and δ. Hence, by (11·51), $b_{\mu\epsilon\nu\sigma} = 0$ unless $E_{\mu\epsilon}$ and $E_{\nu\sigma}$ are both time-like. The dual R.C. tensor thus has at most 36 components. We have

$$\begin{aligned} U_{\alpha\gamma}V_{\alpha\gamma} &= \Sigma\, b_{\mu\epsilon\nu\sigma}(E_{\mu\epsilon})_{\alpha\gamma}(E_{\nu\sigma})_{\alpha\gamma} = \Sigma\, b_{\mu\epsilon\nu\sigma}\,\mathrm{spur}\,(E_{\mu\epsilon}\bar{E}_{\nu\sigma}) \\ &= 4\Sigma\, b_{\mu\epsilon\mu\epsilon} = 4Y, \end{aligned} \tag{11·553}$$

where

$$Y = b_{2323} + b_{3131} + b_{1212} + b_{0404} + b_{0505} + b_{4545}. \tag{11·554}$$

Also we have

$$\begin{aligned} (E_{45})_{\epsilon\zeta}U_{\epsilon\zeta}(E_{45})_{\alpha\gamma}V_{\alpha\gamma} &= \Sigma\, b_{\mu\epsilon\nu\sigma}\,\mathrm{spur}\,(\bar{E}_{45}E_{\mu\epsilon})\,\mathrm{spur}\,(\bar{E}_{45}E_{\nu\sigma}) \\ &= 16 b_{4545}. \end{aligned}$$

Hence, setting $\beta, \delta = \alpha, \gamma$ in (11·552),

$$\tfrac{1}{16}(Q_{\alpha\gamma} - Q_{\gamma\alpha})V_{\alpha\gamma} = Y - 2b_{4545}. \tag{11·56}$$

The term b_{4545} is distinguished by the fact that E_{45} is the product of the two time-like matrices in the pentad $E_{0\sigma}$. This rule enables us to write down the corresponding term for other pentads.

The left-hand side of (11·56) is the quantity which was treated in § 11·4, yielding the equation (11·451). In the symmetrical case it vanishes identically. The right-hand side of (11·56) therefore gives the additional terms which must be included in (11·451) when U and V are antisymmetrical. In place of (11·462), we have the following equations for the diagonal components of $B_{\mu\epsilon\nu\sigma}$

$$\begin{aligned}
&-G_{11}+B_{1515}+B_{1010}+2b_{2323}=B_{1616}+Y \qquad (11\cdot57)\\
&-G_{22}+B_{2525}+B_{2020}+2b_{3131}=B_{1616}+Y\\
&-G_{33}+B_{3535}+B_{3030}+2b_{1212}=B_{1616}+Y\\
&-G_{44}+B_{4545}+B_{4040}+2b_{0505}=B_{1616}+Y\\
&-G_{55}+B_{5050}\qquad\quad+2b_{0404}=B_{1616}+Y\\
&-G_{00}+B_{0505}\qquad\quad+2b_{4545}=B_{1616}+Y
\end{aligned}$$

Adding the first four and subtracting the last two, we have

$$G-2B_{0505}+2Y-4b_{0404}-4b_{4545}=2\,(B_{1616}+Y).$$

So that, setting
$$B_{1616}+Y=\tfrac{1}{2}\,(G-2\lambda), \qquad (11\cdot581)$$

we obtain
$$\lambda=B_{0505}+2b_{0404}+2b_{4545}-Y. \qquad (11\cdot582)$$

The result differs from the symmetrical case in two ways: (1) the value of λ is changed, and (2) the energy tensor $G_{\mu\nu}-\frac{1}{2}g_{\mu\nu}(G-2\lambda)$ now has three constituents. If
$$T_{\mu\nu}=(T_{\mu\nu})_k+(T_{\mu\nu})_e+(T_{\mu\nu})_a, \qquad (11\cdot583)$$

the four diagonal elements of the new constituent are

$$(T_{\mu\mu})_a=(-1/4\pi)\,(b_{2323},\, b_{3131},\, b_{1212},\, b_{0505}). \qquad (11\cdot584)$$

The non-diagonal elements will contain the non-diagonal terms of $b_{\mu\epsilon\nu\sigma}$.

We return now to the dual R.C. tensor with a single component $b_{4545}=b=4R^{-2}$, which gives de Sitter space-time. Then (11·584) vanishes, and the energy tensor consists of kinematical components $B_{\mu5\mu5}$ and electrical components $B_{\mu0\nu0}$ as in the symmetrical case. Here we have $B_{\mu5\mu5}=R^{-2}$ and $B_{\mu0\mu0}=-R^2$, by (11·533); so that $T_{\mu\nu}=0$, as it should be in a de Sitter "empty" world. By (11·582), $\lambda=3R^{-2}$, which agrees with the usual result.†

The "emptiness" of the de Sitter world is here represented as due to a cancelling of two energy tensors $(T_{\mu\nu})_k$ and $(T_{\mu\nu})_e$. This is because the simplest kind of space is an electrically saturated space. To obtain a metric agreeing with that of neutral space we have to cancel the electrical energy. It is therefore appropriate that the formulae should exhibit the zero energy as the result of cancellation. Another way of obtaining the de Sitter world is to take a dual R.C. tensor with two components $b_{4545}=b_{0404}=\frac{1}{2}b$. Since the suffixes 0 and 5 are interchanged in the two components, the four-dimen-

† *Mathematical Theory of Relativity*, equation (69·12).

sional part of $B_{\mu\epsilon\nu\sigma}$ is unaltered; but the energies $B_{\mu5\mu5}$, $B_{\mu0\mu0}$ are found to be zero.

An Einstein world is obtained by taking a dual R.C. tensor with two components

$$b_{4545} = b_{0505} = 2R^{-2}. \tag{11·591}$$

We then find $B_{2323}, B_{3131}, B_{1212}, B_{1515}, B_{2525}, B_{3535} = R^{-2}$,

$$B_{0404}, B_{1616} = -R^{-2}.$$

The other components vanish. These agree with the values in a space-time spherical in three dimensions and cylindrical in the fourth, namely B_{2323}, B_{3131}, $B_{1212} = R^{-2}$; B_{1414}, B_{2424}, $B_{3434} = 0$. For the energy tensor we obtain (remembering that there is a component $(T_{44})_a = -(1/4\pi)\, b_{0505}$)

$$8\pi T_{\mu\mu} = (-1, -1, -1, -3)\, R^{-2}. \tag{11·592}$$

By (11·582), $\lambda = 0$. This is a possible specification of an Einstein world, though it is not the one which has usually been given in general relativity theory. The energy tensor ordinarily adopted is†

$$\begin{aligned} 8\pi T_{\mu\mu}{}' &= (0, 0, 0, -2)\, R^{-2} \\ &= 8\pi T_{\mu\mu} - \lambda' g_{\mu\mu}, \end{aligned} \tag{11·593}$$

where $\lambda' = R^{-2}$. That is to say, (11·593) is obtained from (11·592) by changing the constant λ (the cosmical constant) in the formula

$$-8\pi T_{\mu\nu} = G_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu}(G - 2\lambda).$$

In other words we reckon energy and pressure from a different zero. This change of zero reckoning will be explained in § 14·3.

11·6. Metric and "a priori" Probability.

Let

$$8\pi (T_{\mu\nu})_0 = -(G_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} G), \tag{11·611}$$

$$8\pi (T_{\mu\nu})_c = \lambda g_{\mu\nu}, \tag{11·612}$$

so that

$$T_{\mu\nu} = (T_{\mu\nu})_0 - (T_{\mu\nu})_c. \tag{11·613}$$

We shall suppose tentatively that $(T_{\mu\nu})_0$ and $(T_{\mu\nu})_c$ are the absolute energy tensors of the object system and comparison fluid referred to the frame.‡ The ordinary relative energy tensor $T_{\mu\nu}$ of the object system is thus the difference of the two absolute tensors.

In macroscopic relativity theory $g_{\mu\nu}$ is defined by the condition that $g_{\mu\nu} dx_\mu dx_\nu\, (= ds^2)$ is a quantity measurable in a definite way by scales and clocks. This definition is clearly unsuitable for microscopic theory. We must adopt a definition of $g_{\mu\nu}$ in terms of more primitive conceptions, and show later, when the theory is sufficiently developed to treat the extension of macroscopic aggregations (scales and clocks), that the macroscopic definition is equivalent. We therefore regard (11·612) as the definition of $g_{\mu\nu}$.

† For *real* time the last component becomes $8\pi T_{44}{}' = +2R^{-2}$, giving $\rho = 1/4\pi R^2$.

‡ The absolute energy tensor referred to the frame (self-energy tensor) is defined in § 10·5.

Apart from a numerical scale factor $\lambda/8\pi$, *the metrical tensor* $g_{\mu\nu}$ *is simply the energy tensor of the adopted comparison fluid.*

We have seen that the geometrical frame of coordinates must be filled with a comparison fluid in order to form an ideally observable background of reference for the object systems that we study. The comparison fluid consists of a probability distribution of particles which might be specified in detail; but macroscopically it is sufficiently described by specifying its energy tensor (relative to the frame) as a function of the coordinates. This energy tensor constitutes a metrical tensor for the frame and is more familiar to us under that name. Having thus defined $g_{\mu\nu}$, we can transform the coordinates locally to rectangular coordinates. Then (11·612) becomes, on raising a suffix,

$$(T_\mu{}^\nu)_c = (\lambda/8\pi)\,\delta_\mu{}^\nu. \qquad (11\cdot614)$$

That is to say, the comparison fluid turns out to be uniform and isotropic at every point—so far as the gross characteristics described by the mixed energy tensor are concerned. We do not *choose* an isotropic comparison fluid. Whatever comparison fluid is chosen turns out to be isotropic, because, being the comparison fluid, it is the standard of isotropy.

In elementary problems we consider only one or two particles. It is not intended that there shall be no other particles in the universe. It is implicitly assumed that there is some innocuous way of distributing the other particles so that the results of our investigation will not be entirely invalidated by their presence. Formally what is known as "the problem of two particles" should be described as the problem of N particles, of which two are studied in detail whilst the remaining $N-2$ are supposed to have some standard average distribution which remains undisturbed by the behaviour of the two particles. We shall call the $N-2$ particles *unspecified particles*. A particle is *specified* by giving it a wave function describing a particular probability distribution of position or momentum; so long as its probability distribution is the standard average distribution there is no need to mention it in the problem.

It would be redundant to surround the object system with a double environment of (*a*) unspecified particles waiting to be introduced into the analysis as the treatment grows more comprehensive, and (*b*) the particles constituting the comparison fluid. We therefore adopt a comparison fluid consisting of all the unspecified particles in the universe. Let the number be N. It is convenient also to define a *partial comparison fluid* having $1/N$th of the density of the actual comparison fluid, and therefore corresponding to the standard probability distribution of one unspecified particle. The absolute energy tensor of a partial comparison fluid is, by (11·612),

$$(T_{\mu\nu})_p = (\lambda/8\pi N)\,g_{\mu\nu}. \qquad (11\cdot615)$$

Whilst it is permissible to speak of the energy tensor $T_{\mu\nu}$ belonging to a

particular object particle, we must not use the formulae equating it to curvature of space-time, unless the object particle is the sole occupant of the region considered. Since the probability distributions of a number of object particles may overlap, we must write (11·611) more generally as

$$8\pi\Sigma(T_{\mu\nu})_0 = -(G_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}G). \qquad (11\cdot621)$$

Here Σ extends over all the particles in the universe treated in turn as object particles, although in practice the main contribution comes from a few particles specified as being present in the region we are considering. (The combined density in any region of all the unspecified particles is probably less than 10^{-28} gm. per cu. cm.)

We have therefore for the whole energy tensor in any region

$$T_{\mu\nu} = \Sigma(T_{\mu\nu})_0 - (T_{\mu\nu})_c = \Sigma\{(T_{\mu\nu})_0 - (T_{\mu\nu})_p\}, \qquad (11\cdot622)$$

there being N terms in the summation. Thus each object particle has the relative energy tensor

$$(T_{\mu\nu})_0 - (T_{\mu\nu})_p. \qquad (11\cdot623)$$

When an object particle is unspecified, it has the standard probability distribution, and its absolute energy tensor $(T_{\mu\nu})_0$ is equal to $(T_{\mu\nu})_p$. Thus its relative energy tensor vanishes. It follows that for a specified object particle:

The relative (observable) energy tensor is measured from the standard probability distribution as zero, and not from a hypothetical state of non-existence. The conception of creating a particle does not enter into our theory. *Energy is furnished to a particle by specifying it, not by creating it.*

The result (11·623) is the justification of the tentative procedure which we have been following in this section. It is necessary in order to conform to the general outlook of wave mechanics, which allows us to discuss a system of a few particles without explicit recognition of the vast number of other particles in the universe. If $T_{\mu\nu}$ were merely a tensor involved in the inner mechanics of the system, it would not matter much from what zero it was reckoned. But in unified theory $T_{\mu\nu}$ is to be identified with the relativity tensor $(-1/8\pi)\{G_{\mu\nu} - \frac{1}{2}g_{\mu\nu}(G - 2\lambda)\}$; and for this it is essential that all the particles in the region, specified or not, shall be included. If in the formulation of the "problem of two particles" it is intended that there shall be no unspecified particles overlapping the system, this should be stated explicitly; the appropriate terms must then be introduced into the equation to represent the energy employed in creating a vacant space among the $N - 2$ unspecified particles; in fact the unspecified particles must be *specified as absent* from the region. Since this is not the procedure followed, we must adopt the alternative given by (11·623), namely we must choose the zero of energy in such a way that the unspecified particles make no contribution to $T_{\mu\nu}$ so that we may ignore them without harm.

Our results up to this point are:

(*a*) The adopted comparison fluid consists of the unspecified particles of the universe.

(*b*) The absolute energy tensor of the comparison fluid is the λ-term (generally known as the cosmical term) in the relative energy tensor.

(*c*) The absolute energy tensor of the comparison fluid is the metrical tensor of the frame.

If a different comparison fluid were used it would be necessary to "specify", either individually or macroscopically, the particles composing it. The macroscopic reference objects used in particular observations can be regarded as a *specified comparison fluid.* But it is simpler to treat these as additional object systems; so that the same comparison fluid (*a*) is used as intermediary in all problems, and the observed relations between object systems are resolved into relations (unobserved, but nevertheless ideally observable) between each object system and a common comparison fluid.

The probability distribution which applies in the absence of special information is commonly called the *a priori* probability distribution. The unspecified particles are those as to which we have no special information. The probability distribution of an unspecified particle, which we have identified with a partial comparison fluid, is therefore the *a priori* probability distribution. The identification

A priori probability distribution = partial comparison fluid

helps to connect our theory with current terminology; but it must be understood that we reject emphatically the *conception* of *a priori* probability. The distribution is uniform and isotropic not for *a priori* reasons, but for the reason already given, namely that it determines the tensor $g_{\mu\nu}$. It is, I think, generally admitted that the need for a conception of *a priori* probability is a logical weakness in statistical mechanics as ordinarily developed. We need not discuss here attempts to defend it on a metaphysical basis, and to deduce from a "Principle of Indifference" that it has a uniform and isotropic distribution. Probably few theorists would accept such views today. But it is not generally realised that relativity theory has rendered the conception of *a priori* probability entirely unnecessary.†

The so-called *a priori* probability distribution used in statistical mechanics is essentially an unobservable; for as soon as we have any observational information, the *a priori* probability ceases to apply, and a modified (actual) probability is substituted. *A priori* probability is to be treated therefore as other unobservables (introduced for mathematical convenience) are treated in relativity theory, e.g. frames of space and time. The *a priori* probability distribution is sometimes called a "basis of statistics"; that is to say, it is

† *New Pathways in Science,* pp. 129ff.

a standard of reference for statistical enumerations in the same way that a frame of space and time is a standard of reference for measurements of extension. Observable results are invariant for all transformations of these reference frames; they must equally be invariant for changes of the adopted *a priori* probability distribution. The common appellation is therefore a misnomer; the *a priori* probability distribution is not anything that is given *a priori*; it is an arbitrary comparison distribution which we can change as we please.

The reason why it is not ordinarily recognised that observable phenomena are invariant for transformations of the *a priori* probability distribution, is that the transformation is not pursued to the end. If the change of reference distribution changes the equations of an electron and proton, it will change also the equations determining the behaviour of the measuring appliances composed of electrons and protons. When the transformation is applied both to the measuring appliance and the object measured, the result of the measurement is invariant. But if we apply the transformation only to the microscopic systems which are being measured and not to the macroscopic measuring apparatus, we fail to notice this invariance. That was the mistake originally made in interpreting the Michelson-Morley experiment; the transformation representing change of motion was at first applied only to the optical factors in the experiment, the mechanical or metrical factors being overlooked.

This invariance is provided for in the usual way by a tensor relation between the quantities concerned in the microscopic and macroscopic parts of the theory. In the microscopic equations the change of *a priori* probability is formulated as a change of the energy tensor $(T_{\mu\nu})_c$ of the comparison fluid; in the macroscopic equations (which determine the behaviour of the measuring appliances) the same change is expressed as a change of the metrical tensor $g_{\mu\nu}$.

The linkage of the changes of $g_{\mu\nu}$ and $(T_{\mu\nu})_c$, so as to produce no observable result, is provided for by their description as tensors of the same kind. In (11·612) we have gone further, and identified them. The importance of this step is that if the frame and the comparison fluid are both described by the same tensor—called in one case the metrical tensor and the other case the energy tensor—it minimises the error of current quantum theory in confusing them. To give meaning to the ordinary equations of current quantum theory we have to suppose that a particular physical reference object is intended, and we have to discover what this reference object is. If, on being challenged, the quantum physicist were to reply that he intended the reference object to be an atom of carbon travelling at 150 km. per sec. in the x-direction, it would be preposterously misleading not to have stated this. If, however, he replies that he has not distinguished between the frame

and the comparison fluid, so that when he said that the tensor $g_{\mu\nu}$ of the frame was isotropic (i.e. that his coordinates were rectangular) he also meant that the comparison fluid was isotropic, his omission is a minor inadvertence. He might perhaps accuse us of stupidity if we did not supply the omission for ourselves. Having regard to this, our identification of $g_{\mu\nu}$ with the energy tensor of the comparison fluid is an obvious rectification of an omission in current quantum theory.

Locally any change of $(T_{\mu\nu})_c$ or $g_{\mu\nu}$ is a tensor transformation; but over an extended region we have to distinguish between tensor transformations and absolute changes. From the second derivatives of the $g_{\mu\nu}$ we obtain certain characteristics of the comparison fluid, which we shall call hypercharacteristics, which must be specified definitely in any problem and must not be changed in any transformation. These hypercharacteristics correspond to irreducible gravitational fields, or to the distributions of matter causing such fields. Either an assurance that the field is zero, or a specification of the invariants of the field, is an essential datum of any problem concerning an object system which covers an extended region. The observable results are not invariant for transformations of the probability distribution of the unspecified particles which would create a change of irreducible gravitational field; and when, as usual, the comparison fluid is not mentioned explicitly as a datum of the problem, it is obviously implied that its characteristics are not incompatible with the other data furnished, that is to say with the invariants of the gravitational field.

Accordingly the statement that any distribution may be chosen as an *a priori* probability distribution requires qualification when an extended region is considered. It must be remembered that we are not defending *a priori* probability. In our theory the partial comparison fluid takes its place. Therefore we are not led into a discussion as to what kind of field should be assumed as probable *a priori*, if no observational information is furnished. If we are not told what the field is, we have no idea what will happen, or what is likely to happen—and there is no more to be said.

By (11·623) the unspecified particles of the universe make no contribution to $T_{\mu\nu}$, and can therefore be neglected in elementary problems—at any rate so far as this part of the theory is concerned. But there are other ways in which the treatment of a system of a few particles as if they were the only particles in the universe is liable to be misleading. It is natural to discuss the effect of spatial displacement of a single particle, the other particles in the universe being unchanged. It is possible also to discuss a temporal displacement of a single particle, the other particles being unchanged. This conception is used frequently and legitimately in relativity theory; indeed it would be contrary to the fundamental principle of relativity theory to discriminate between space and time displacements. But in quantum

physics the notion of temporal displacement has been employed less carefully. Usually when reference is made to a displacement (dx, dy, dz, dt) of a particle, it seems to be intended that (dx, dy, dz) applies to that particle, and dt applies to all the particles in the universe. Obviously, when this interpretation is intended, Lorentz transformations are not applicable to (dx, dy, dz, dt). Thus the common form of the "problem of one particle" combines spatial displacement of one particle with temporal displacement of N particles. The "one" particle is thus a hybrid whose time displacements are N times as effective as its space displacements in producing recoil of the comparison fluid; and the corresponding Riemann-Christoffel tensor has its time component B_{4545} magnified N times relatively to its space components B_{1515}, etc. The corresponding radius of curvature is therefore R in the space directions, but $R' = R/\sqrt{N}$ in the time direction. Since N is about 10^{79}, the geometrical picture has become too extravagant to be retained; but we express the same analytical relation by ascribing to the intrinsic vectors of the object particle the appropriate rotation $\delta\theta'$, linked by B_{4545} to the rotation $\delta t/R$ of its position vector in space-time. This very rapid rotation of the intrinsic vectors (e.g. the stream vector) is interpreted as a periodic wave. The wave length should be $2\pi R'$, which is of order 10^{-12} cm., and is comparable with the wave length of the de Broglie waves of elementary particles. The precise calculation of the wave lengths for protons and electrons involves additional considerations which will be treated in due course.

The waves attributed to an electron in current wave mechanics belong to it, not individually, but because it is one of N particles; and their wave length is dependent on the number N. Considered by itself, a particle has no other periodicity than that due to the connectivity of the space in which it is situated. The wave length of its "waves" is the circumference of space.

11·7. Microscopic and Macroscopic Theory.

In the most general formulation of the problems of physics the particles are divided into three groups:

(a) One or more particles specified individually.

(b) Collections of particles specified macroscopically.

(c) The "rest of the universe" consisting of unspecified particles.

In mathematical treatment the subject is divided into a number of idealised problems in which one of the groups may be absent.

In general the particles of group (c) far outnumber all the others, and most of our formulae are developed on this understanding. There is, however, a class of problems, namely, cosmical problems, in which the whole universe is specified macroscopically, e.g. an Einstein universe; there are then no

particles of group (*c*), and special precautions must be taken in applying the formulae (Chapter XIV).

By (11·623) the energy tensor of the unspecified particles vanishes. But there is no such condition for the other components of their R.C. matrix; and the four-dimensional R.C. tensor of the unspecified particles is the foundation of the R.C. tensor of space-time. We shall call this the *cosmical R.C. tensor*. If it exists alone it gives a de Sitter space-time; for that is well known to be the only solution (without singularities) of the equation $G_{\mu\nu} - \frac{1}{2}g_{\mu\nu}(G - 2\lambda) = 0$ satisfied by space-time containing only unspecified particles.

Notwithstanding the overwhelming preponderance of unspecified particles, the effect of group (*b*) is often more important. This is because the effect is more localised. As part of their macroscopic specification, the *position* of these collections of particles is closely defined; and they contribute an R.C. tensor which locally far outweighs the cosmical R.C. tensor. Consequently the latter is commonly neglected in the local problems of general relativity theory, e.g. the theory of planetary motion.

Quantum theory deals primarily with problems involving (*a*) and (*c*) only. The influence of macroscopic objects, if any, on these problems is attributed to gravitational or electromagnetic fields emanating from them. We may therefore call problems which do not involve group (*b*), directly or indirectly, *field-free problems*. In field-free conditions space-time is of the de Sitter type. The specified particles (*a*) are represented by wave-functions, and the unspecified particles (*c*) are represented by the cosmical R.C. tensor, or equivalently by the $g_{\mu\nu}$ which correspond to it.

It might be thought that the individually specified particles, like those of group (*b*), would contribute a local R.C. tensor very much greater than the cosmical tensor. For example, the average density in the space occupied by an atom is very much greater than the density ($\sim 10^{-28}$) due to unspecified particles. But this is a confusion of ideas. We must make up our minds whether the R.C. matrix of a particle is to be absorbed into the R.C. tensor of space-time, or reserved for more detailed treatment by quantum theory. *We cannot have it both ways*. A particle specified individually is one of those reserved for detailed treatment. Its interaction with its surroundings is expressed by wave functions. We must not at the same time insert a duplicate in the R.C. tensor, and so obtain another (cruder) representation of its interaction by the changes of $g_{\mu\nu}$ which would result.

We shall find later that the wave functions of a specified particle, employed in current theory, are defined in such a way that they represent an addition to, not a substitution for, the characteristics belonging to it as an unspecified particle. In that sense the specified particle is represented in the R.C. tensor as well as in the wave functions. But that merely secures that the cosmical

tensor is not disturbed when the particle is specified. If it were not for this provision, the R.C. tensor of space-time would be reduced by about 1 part in 10^{79} for every particle specified. As it is, we may regard the number of unspecified particles as a fixed constant N, giving a constant R.C. tensor however many particles may be specified individually.

It is important to realise that the individual atoms and electrons, whose energy and momentum are given by their wave functions, do not disturb the curvature of the space-time in which their wave functions are represented. Reciprocally the macroscopic objects, whose energy and momentum are represented by components of curvature of space-time, have no wave functions.† For the purpose of investigating the connection of microscopic and macroscopic theories, we can represent macroscopic objects by complicated wave functions; but such wave functions occupy a space-time which has no more than the cosmical curvature. In § 11·1 we referred to the error of confusing the frame with the physical comparison system, and remarked that the neglect of uncertainty of position and momentum of light comparison objects and the neglect of curvature of space-time produced by heavy comparison objects were different forms of the same error. The error takes one form or the other according as the comparison objects are specified microscopically or macroscopically.

Using the notation of § 10·4, the R.C. matrix can be expressed in the various forms:

$$(B)_{\alpha\gamma\beta\delta} = \Sigma \Psi_{\alpha\beta} \mathrm{X}_{\gamma\delta} = \Sigma U_{\alpha\gamma} V_{\beta\delta} \tag{11·711}$$

$$= \Sigma B_{\mu\epsilon\nu\sigma} C_{\mu\epsilon} D_{\nu\sigma} = \Sigma b_{\mu\epsilon\nu\sigma} E_{\mu\epsilon} F_{\nu\sigma} \tag{11·712}$$

$$= \Sigma B_{\mu\epsilon\nu\sigma} (E_{\mu\epsilon})_{\alpha\beta} (E_{\nu\sigma})_{\gamma\delta} = \Sigma b_{\mu\epsilon\nu\sigma} (E_{\mu\epsilon})_{\alpha\gamma} (E_{\nu\sigma})_{\beta\delta}. \tag{11·713}$$

Let us consider in particular the cosmical R.C. matrix which gives a metric of the de Sitter form. By (11·533) this is‡

$$(B) = 4R^{-2} E_{45} F_{45} = R^{-2} \{\Sigma_k C_\sigma D_\sigma - \Sigma_e C_\sigma D_\sigma\}. \tag{11·721}$$

This is the space tensor; the corresponding strain tensor is

$$(B)_{\text{strain}} = 4R^{-2} E_{16} F_{16} = R^{-2} \{\Sigma_s C_\sigma D_\sigma - \Sigma_t C_\sigma D_\sigma\} \tag{11·722}$$

as in (10·54). By § 10·5 this represents an "uncertain" momentum in the CD-frame, or an "exact" momentum in the EF-frame.

Consider first the CD-frame. By (11·712) the R.C. tensor $B_{\mu\epsilon\nu\sigma}$, which constitutes the complete energy tensor, has been constructed in the CD-frame; and it would therefore seem that the uncertain energy tensor is the true cosmical energy tensor. But in that case the position tensor is exact;

† Or we may use wave functions which represent the kinetic energy and pressure of the particles of a gas, but not their rest energy; in that case the curvature of the space-time in which these functions are represented corresponds to the rest energy, but not the kinetic energy.

‡ The suffixes k, e, s, t refer to kinematical, electrical, space-like and time-like matrices.

that is to say the cosmical R.C. tensor is wholly concentrated at one point of space. This, of course, does not correspond to the actual conditions of de Sitter space-time. The failure is easily explained; we cannot do more than we have done with one double frame. The cosmical R.C. tensor is made up of contributions from some 10^{79} unspecified particles, each of which has its own double frame. If we attribute to them uncertain momenta, they will have exact positions; but between them they will cover all space, giving a practically continuous R.C. tensor.

Since $(B)_{\text{strain}}$ is algebraic in the EF-frame, we can write, instead of (11·722),

$$(B)_{\text{strain}} = -4R^{-2} = (4/R^2N)\{E_{16}F_{16} + E_{16}'F_{16}' + \ldots + E_{16}^{(N)}F_{16}^{(N)}\}, \tag{11·73}$$

introducing N double frames instead of one. If the frames commute (Case (b) in § 2·7) they have no relation of orientation to one another. The method of imposing a uniform distribution of orientation belongs to the type of treatment of the many particle problem introduced in Chapter XVI. Here it is sufficient to recognise that the difficulty of avoiding concentration of the R.C. tensor at one point will disappear when we take account of its highly composite nature.

But we have a great deal more to learn from the double frame, before proceeding to the many particle problem. We consider the simplest of all problems—an elementary object particle in de Sitter space-time. The unspecified particles, which provide the de Sitter metric, form the comparison fluid for the particle. To bring the problem within range of treatment by double wave functions, we use a partial comparison fluid consisting of one unspecified particle. The object particle, being specified microscopically, does not contribute to the curvature of the space-time in which we represent its wave function. The cosmical R.C. tensor at any point corresponds to a uniform distribution of momentum vectors in all directions. We could suppose that this was a superposition of exact momentum vectors of a great number of unspecified particles; but it is evidently more convenient to take it to be the entirely uncertain momentum of one unspecified particle—a partial comparison fluid. This will have definite position in the small region we are considering; but the other unspecified particles will provide similar R.C. tensors at other points. (We take advantage of our liberty to choose any convenient way of analysing the unspecified matter of the universe into "particles", i.e. wave systems.) The cosmical R.C. tensor at the point is the complete self-energy tensor of the partial comparison fluid. We have represented it in a CD-frame. By transforming to the EF-frame (so that the R.C. tensor is transformed into its dual $b_{\mu\epsilon\nu\sigma}$) we change it into an exact energy tensor $bE_{45}F_{45}$. We then readily recognise it as the

outer square of the stream vector of a neutral particle which was identified with the self energy tensor in § 10·5. The unspecified particle is neutral, for otherwise it would not give the R.C. tensor characteristic of neutral space-time; also when the "entire uncertainty" applies to all components of a complete stream vector the sign of the charge necessarily shares in the uncertainty.

Turning to the elementary object particle, since we are considering only a point, or small region, we must for consistency suppose it to have exact position and uncertain momentum. This condition is not easy to formulate since we do not want the charge and spin to be uncertain. It is simplest to start with an elementary particle and its partial comparison fluid defined as having exact momentum in the EF-frame. We can afterwards transpose the results to the CD-frame. The product of the stream vectors of the object particle and of the partial comparison fluid will give the mutual energy tensor. This when transposed to the CD-frame will give the R.C. matrix of the object particle. The portion $B_{\mu 5 \nu 5} + B_{\mu 0 \nu 0}$ will give the ordinary energy tensor of the object particle. This as we have seen is the observable relative energy. Thus the *relative energy* defined by $T_{\mu\nu}$ is the same as the *mutual energy* defined in § 10·5 as the product of the stream vectors; except that the latter is expressed as a complete tensor and exhibits the kinematical and electrical energies separately, whereas $T_{\mu\nu}$ gives their sum. We can separate the part corresponding to the self energy of the comparison fluid, as in (11·613), the remainder being regarded as an energy of the object particle.† For consistency this should agree with the self energy of the object particle defined directly in (10·51). This condition is investigated in § 12·6.

11·8. Neutral Comparison Fluid.

By working in the EF-frame we are able to use particles with exact stream vectors. Consider an object particle with stream vector $U = \Sigma E_\mu u_\mu = \psi\chi^*$ and a comparison particle with stream vector $V = \Sigma E_\mu v_\mu = \phi\omega^*$. The combined system is described by double wave vectors Ψ, X, where

$$\Psi = \psi\phi^*, \quad \mathrm{X} = \chi\omega^*. \tag{11·81}$$

The microscopic treatment will depend on the use of these wave functions; but the connection with macroscopic theory is given by the identification

† Since the object particle and comparison fluid move in opposite directions relative to the frame in our coordinates (space and imaginary time), the self energy of the partial comparison fluid should be taken to be *minus* the square of the stream vector in order that it may be comparable with a mutual energy. An unspecified particle treated as object particle would produce a recoil of a similar unspecified particle treated as its partial comparison fluid. This explains why the total energy in (11·613) is the difference, not the sum, of the energies of the object particle and comparison fluid.

(comparing (11·711) and (11·713))

$$u_{\mu\epsilon}v_{\nu\sigma}=b_{\mu\epsilon\nu\sigma}, \tag{11·821}$$

where $b_{\mu\epsilon\nu\sigma}$ is the dual R.C. tensor of the object particle with the other particle as comparison fluid, together with the condition that the dual cosmical R.C. tensor $b_{\mu\epsilon\nu\sigma}{}'$ which determines the metric is given by

$$v_{\mu\epsilon}v_{\nu\sigma}=b_{\mu\epsilon\nu\sigma}{}'. \tag{11·822}$$

We have been mainly using space tensors, because we wished to connect our results with the space tensors of general relativity theory; but for the special purposes of wave mechanics strain vectors are more useful. We shall therefore take ψ, χ, ϕ, ω to be covariant wave vectors forming a strain tensor $b_{\mu\epsilon\nu\sigma}$.

In this simple formulation we have taken the comparison particle (or partial comparison fluid) to be an elementary charged particle. Our standard equations, however, postulate a neutral comparison fluid. We wish now to direct special attention to the step by which we pass from a charged particle to a neutral particle as comparison fluid. The relation between the stream strain vectors V, V_0 of an elementary and a neutral particle is

$$V_0=\mathrm{qs}\,V. \tag{11·83}$$

This follows at once from (6·64), the strain vector of the neutral particle (equally likely to be in any one of the four states) being $\frac{1}{4}(S_a+S_b+S_c+S_d)$.

We can therefore pass from an electrically saturated to a neutral comparison fluid by contracting the suffixes of ϕ and ω. The change is expressed by the transformation

$$\psi_\alpha\phi_\beta\chi_\gamma\omega_\delta\rightarrow\tfrac{1}{4}\psi_\alpha\phi_\beta\chi_\gamma\omega_\beta \tag{11·84}$$

or, in terms of wave functions,

$$\Psi_{\alpha\beta}\mathrm{X}_{\gamma\delta}\rightarrow\tfrac{1}{4}\Psi_{\alpha\beta}\mathrm{X}_{\gamma\beta}. \tag{11·85}$$

In (11·85) we need no longer assume that Ψ and X are factorisable. A state of a double system is regarded as elementary if it is pure as regards factorisation into Ψ and X, irrespective of whether there is a UV factorisation. The formula (11·85) is of great importance in Chapter XII.

We have at last reached the solution of what might be called the initial problem of electron and proton theory. The microscopic complete relative (or mutual) energy strain tensor of an elementary charged particle together with a neutral comparison fluid has the form $\Psi_{\alpha\beta}\mathrm{X}_{\gamma\beta}$. As the importance of this may not be obvious, we make the following comments:

(1) The elementary charged particle is usually described by simple wave functions ψ, χ, and the energy and momentum are contained in a simple stream vector $\psi\chi^*$. By introducing an empirical mass constant (different for the proton and electron) these simple wave functions are made to serve the ordinary purposes of quantum theory and yield correct results. But if

we are not content to accept the mass as an empirical characteristic peculiar to wave mechanics, and wish to trace its connection with the macroscopic mass identified with a component of space-time curvature, we are met with the difficulty that all the measurable characteristics of a charged particle are contained in *double* wave functions of the particle and a physical comparison system. The conditions of replacement of the double wave functions by single wave functions is the central problem in the unification of relativity theory and current quantum theory. We take a large step towards the solution when we find that, adopting a neutral comparison fluid, the double wave functions are of the specialised kind $\Psi_{\alpha\beta}$, $\mathrm{X}_{\gamma\beta}$, the second suffixes being "contractible".

(2) Unlike ψ and χ, the wave functions Ψ, X are directly connected with the fundamental tensors of relativity theory. Taking the energy tensors of a number of particles (such that the components ignored in macroscopic theory cancel out) we can calculate the curvature of space-time equivalent to them, and hence translate the energy, etc., into macroscopic units.

(3) The replacement of the energy tensor $\Psi\mathrm{X}^*$ by an energy vector $\psi\chi^*$ occurs by contraction of the suffixes β, δ, which is the usual method in tensor calculus of reducing the rank of a tensor.

(4) We have taken the comparison fluid to consist of one neutral particle, that being all that we can provide for with one double frame. It is therefore a partial comparison fluid according to our previous terminology. The passage to the actual comparison fluid consisting of N unspecified particles is, however, quite simple. The analysis in (11·73) represents the energy tensor as the sum of independent tensors in different frames; so that each object particle affects only its own partial comparison fluid. The average recoil of the total comparison fluid is therefore $1/N$th of the amount of the recoil of the partial comparison fluid. The Riemann-Christoffel tensor is thereby lessened in the ratio $1/N$.

(5) The ideal conditions postulated in these equations correspond to de Sitter space-time containing no macroscopically specified matter. It should be remembered that de Sitter space-time is, according to ordinary standards, an expanding universe. It will be necessary later to treat the modifications involved in using a static frame of reference (Einstein space-time) but this involves introducing macroscopically specified matter.

We shall now show that the contractible double wave vectors $\Psi_{\alpha\beta}$, $\mathrm{X}_{\gamma\beta}$ can (for practical purposes) be replaced by simple wave vectors ψ_α, χ_γ. The observational significance of wave functions is that they determine expectation values for the system of operators P which represent observable relations. Ordinarily for double wave vectors (appropriately normalised) the expectation value is $\mathrm{X}_{\alpha\beta}P_{\alpha\gamma,\beta\delta}\Psi_{\gamma\delta}$; but in the present application the

contraction of the second suffixes restricts P to the form $P_{\alpha\gamma}$, and the expectation value is

$$X_{\alpha\beta}P_{\alpha\gamma}\Psi_{\gamma\beta}. \tag{11·86}$$

Introduce eight simple wave vectors $\psi^{(\beta)}$, $\chi^{(\beta)}$ $(\beta=1, 2, 3, 4)$ defined by

$$\psi_\alpha{}^{(\beta)}=\Psi_{\alpha\beta}, \quad \chi_\alpha{}^{(\beta)}=X_{\alpha\beta}.$$

Then (11·86) becomes

$$\chi_\alpha{}^{(1)}P_{\alpha\gamma}\psi_\gamma{}^{(1)}+\chi_\alpha{}^{(2)}P_{\alpha\gamma}\psi_\gamma{}^{(2)}+\chi_\alpha{}^{(3)}P_{\alpha\gamma}\psi_\gamma{}^{(3)}+\chi_\alpha{}^{(4)}P_{\alpha\gamma}\psi_\gamma{}^{(4)}. \tag{11·87}$$

This is the expectation value of P with respect to an impure simple wave tensor

$$T=\psi^{(1)}\chi^{(1)}+\psi^{(2)}\chi^{(2)}+\psi^{(3)}\chi^{(3)}+\psi^{(4)}\chi^{(4)}. \tag{11·88}$$

By using symbolic factors of T the expectation value may be written $\Gamma P \Gamma$ as in (3·35).

Thus, for the purpose of calculating expectation values, the pure (contractible) double tensor ΨX^* can be replaced by the impure simple tensor T. The reduction of double wave tensors to simple wave tensors is therefore legitimate, but involves a re-analysis into pure elementary states.

We have considered the reduction of an isolated (discrete) wave tensor. There still remains the question of replacing the phase space of the double wave tensor by that of a simple wave tensor. We shall see in the next chapter that this transformation of the phase space has important consequences.

11·9. Retrospect.

In this chapter a gulf between macroscopic relativity theory and microscopic quantum theory has been bridged. We have found how to construct a Riemann-Christoffel matrix, (*a*) out of the tensors employed in general relativity theory, and (*b*) out of the wave functions employed in quantum theory. The mechanical quantities, energy, momentum, etc. referred to in the two theories are thus reduced to a common form. We can, for example, add the R.C. matrices corresponding to the wave functions of a large number of particles and derive the energy tensor which describes the aggregation macroscopically.

Whereas the construction of the R.C. matrix out of macroscopic tensors is straightforward, the theory of its construction out of wave functions is perhaps unexpectedly intricate. It is natural to put the question, What is the R.C. matrix corresponding to a proton or electron with an exact momentum vector? The question is not so elementary as it seems, because the properties ascribed to the particle are really combined properties of the particle and a standard comparison fluid; and although the particle is elementary, the standard comparison fluid is not. If we consider instead an object particle referred to a single comparison particle, both having exact momentum vectors, the answer is simple. The R.C. matrix

is the outer product of their stream vectors. Already in Chapter x, the portion of this R.C. matrix which specifies linkage of translations, viz. the kinematical energy tensor, has been called the "mutual energy tensor" of the two particles; but this was merely a development of the nomenclature of current quantum theory. The connection with energy as classically, i.e. macroscopically, defined appears for the first time in the present Chapter.

The standard comparison fluid has an isotropy in space and time which can be defined by the condition that its energy tensor (referred to the frame) is invariant for rotations and Lorentz transformations, or equivalently by the condition that its energy tensor† is to be identified with the metrical tensor $g_{\mu\nu}$. To pass from an elementary comparison particle to this standard comparison fluid, it is not sufficient to integrate over a symmetrical distribution of stream vectors of the comparison particle. The symmetry is disturbed by the recoil due to the motion of the object particle (specified by its exact stream vector). To neglect this recoil would be to neglect the very thing we are investigating, namely the mechanical specification of the object particle; for in the idealised universe consisting of an object particle and comparison fluid, the only manifestation of the mass of the object particle is in its mechanical reaction on the comparison fluid.

In treating the construction of the R.C. matrix, and the connection of mass and momentum in quantum theory with the corresponding macroscopic tensors, it is essential to bear in mind that displacement of the object particle disturbs the comparison fluid, the combination of action and reaction being expressed by double wave functions. But in applying this theory to the elementary wave functions of quantum theory, the complication arises that these are adapted to a different point of view. These simple wave functions are intended to describe self-contained systems superposed on an *undisturbed* environment. The conception of such detached systems is essentially non-relativistic, but it can be justified up to a point as a practical procedure. It is evidently necessary that the energy and momentum associated with the simple wave functions should include, not only that which properly belongs to the object particle, but also the recoil energy and momentum communicated to the comparison fluid. We have therefore to separate out from the whole R.C. matrix the constant portion which corresponds to the undisturbed comparison fluid (and provides a constant metric); the remainder is then regarded as the self energy of entities which have been added to the comparison fluid without disturbing it. The actual calculation of this self energy, for an elementary particle added to the standard comparison fluid, will be treated in Chapter xii.

† More precisely, its kinematical self energy tensor.

A much easier problem is that of an object particle with an entirely uncertain momentum vector, whose recoil accordingly does not disturb the symmetry of the comparison fluid. Complete uncertainty involves uncertainty of charge and spin, so that the particle is necessarily neutral. To calculate the corresponding R.C. matrix we project the uncertain stream vectors of the object particle and comparison fluid from the *EF* frame to the *CD* frame. In the latter frame the stream vectors become exact; we therefore form their outer product, and project back into the *EF* frame. The resulting R.C. matrix is of the form (11·721), and gives a R.C. tensor which represents de Sitter space-time.

We can regard the standard comparison fluid as composed of N neutral particles with uncertain momentum vectors, and with exact position vectors uniformly distributed in space and time.† Generally in treating a single object particle we use only one of these particles as comparison fluid—a partial comparison fluid. In any problem relating to the actual universe the N particles of the comparison fluid must be present as object particles, although only a few of them are mentioned explicitly; the remainder are either unspecified or specified macroscopically. It is therefore convenient to pair the particles, and treat each object particle as disturbing only its allotted comparison particle. The substitution of a partial for a total comparison fluid magnifies the recoil, and therefore the apparent mass, in the ratio N; this factor is absorbed into the ratio of the units used respectively in macroscopic theory and quantum theory, the elementary equations and definitions of the latter theory being, as it were, based on the assumption that the particle under consideration is the only particle in the universe. We shall find later that more precisely the factor is $\frac{3}{5}N$; the modification is due to the exclusion effect of the particles of the comparison fluid on one another.

When all the particles are unspecified we obtain the cosmical R.C. matrix. Extracting from it the macroscopic R.C. tensor, we find that it corresponds to de Sitter metric. The energy tensor $T_{\mu\nu}$ vanishes; this is the relative tensor of the aggregation treated as object system referred to the same aggregation treated as comparison fluid. The relative energy tensor is closely related to the mutual energy tensor (which in this case becomes a self energy tensor); but in the mutual or self energy tensor the kinematical and electrical components are kept distinct, and the former constitute the energy components in the more limited sense. The kinematical self-energy $B_{\mu 5 \nu 5}$ does not vanish, and, as already explained, it is identified with $g_{\mu\nu}$.

To obtain any other than a de Sitter metric it is necessary to specify

† This is for analytical convenience; the actual distribution in position is, of course, continuous.

some particles macroscopically. The only important alternative metric is that of an Einstein universe; this will be treated in Chapter XIV. Irregular metrics are important in general relativity theory, but it seems unprofitable to combine them with microscopic problems, e.g. to investigate the behaviour of a hydrogen atom in an irreducible gravitational field. Practically the perturbation would be exceedingly minute; and if we retain such small quantities, it is presumably illegitimate to employ a macroscopic specification of the perturbing system. In any case the problem is concerned with a special kind of perturbation of the atom and it would be out of place to deal with it in general theory. For the purposes of quantum mechanics we are therefore limited to the two possible uniform metrics, viz. the de Sitter and Einstein universes.

The prominence of the de Sitter metric in our investigations is due to the fact that, following Dirac, we have developed the theory from Lorentz-invariant equations. Lorentz transformations are not applicable to an Einstein universe. So long as we treat free particles, with exact momentum vectors and therefore distributed uniformly throughout the universe, a Lorentz-invariant frame of de Sitter type is required. When we come to less abstract problems, and treat steady states, represented by internal wave functions to which Lorentz transformations are inapplicable, an Einstein metric becomes admissible; and it is to be preferred because (unlike the de Sitter metric) it forms a static reference system. Current quantum theory in its more abstract formulae presupposes a de Sitter background, and in its more practical formulae presupposes an Einstein background. Perhaps the most difficult part of the present investigation has been the sorting of these two influences.

CHAPTER XII

THE MASS-RATIO OF THE PROTON AND ELECTRON

12·1. Contraction of a Volume Element.

This chapter deals with a point which arises when the double strain vector,

$$S = UV = \psi_\alpha \phi_\beta \chi_\gamma \omega_\delta, \qquad (12{\cdot}111)$$

of an elementary particle, with another elementary particle as comparison fluid, is contracted to form the simple strain vector,

$$\{S\} = U\{V\} = \psi_\alpha \phi_\beta \chi_\gamma \omega_\beta, \qquad (12{\cdot}112)$$

which represents the same particle with a neutral particle as comparison fluid (11·84).†

We may remind ourselves how this process arises. From an observational point of view, it is meaningless to describe a particle without some physical reference object (§ 11·1); and the observable characteristics which we measure are properties of the combined system of the particle and reference object. The wave functions containing observable characteristics, such as the mass of the particle, must therefore be double wave functions Ψ, X, of the particle and an idealised standard reference object which we call the comparison fluid. But in current quantum theory the mass and other characteristics of the particle are assumed to be contained in simple wave functions ψ, χ. How comes it that double wave functions Ψ, X can be replaced in practice by simple wave functions ψ, χ, apparently without much harm? The first part of the answer was reached in § 11·8, when we found that, by adopting the probability distribution of a *neutral* particle as comparison fluid, the double wave vectors $\Psi = \psi_\alpha \phi_\beta$, $X = \chi_\gamma \omega_\delta$ are contracted with respect to the suffixes β and δ which refer to the comparison fluid; so that, for some purposes at least, we can ignore the second suffix and treat Ψ and X as simple wave vectors. In particular the double strain vector S is contracted to $\{S\}$, and becomes the product (actual or symbolic) of two simple wave vectors ψ, χ, together with a merely algebraic factor $\phi_\beta \omega_\beta$.

It is not our business to supply a complete defence of this substitution of simple wave vectors for double wave vectors. It is a substitution which current theory has inadvertently made; and we have to trace how its consequences appear in the formulae of current theory. For our own part we shall be continually going back to the double wave vectors to see what current theory has missed by this substitution.

† The factorisation in (12·111) and (12·112) need only be symbolic.

The strain vector (12·111) specifies an elementary configuration of the combined system—represented by a single point in its phase space. The actual state is specified by probability factors attached to a number of such configurations, or more generally by a continuous probability distribution over the phase space. In the latter case S is normalised so that $S d\omega/\Omega$ is (apart from a unitary factor containing the phases) the probability of a range of configurations $d\omega$ (§ 7·7).

The question that we have to consider is, What happens to $d\omega$ when we contract (12·111)?

Our procedure is based on the law of multiplication of probabilities. The intention is that, leaving out the unitary phase factors, $S = UV$ shall express the law that the probability of a configuration of the double system described by S is the product of the probabilities of the corresponding configurations of the simple systems described by U and V. But $S = UV$ does not express this unless volume elements are inserted on both sides. The formulae as they stand postulate that the different configurations are represented by discrete wave functions; they can only be extended to continuous wave functions if it is assumed (or arranged) that the volume elements take care of themselves.

It will now be convenient to change the notation. Let S, S', S_0 be the strain vectors of the object system, comparison fluid and combined system;† and let $d\omega$, $d\omega'$, $d\omega_0$ be volume elements in their respective phase spaces. To neutralise the comparison fluid we add together four elements $d\omega_0$ which yield the required balance of charge and spin; and we equate the probability contained in these four elements to the product of the probabilities in the corresponding elements of the phase spaces of S and S'. Denoting the summation of four elements by $\{\ \}$, this gives (disregarding unitary factors)

$$\{S_0\} d\omega_0/\Omega_0 = (S d\omega/\Omega)(S' d\omega'/\Omega'). \qquad (12·12)$$

In order that this may reduce to $\{S_0\} = SS'$ as currently assumed, we must have

$$d\omega_0 = k\, d\omega\, d\omega', \qquad (12·13)$$

where k is the constant $\Omega_0/\Omega\Omega'$.

The point at which the assumption $\{S_0\} = SS'$ enters fundamentally into current theory is in the dynamical equations (§ 8·4). As explained in § 9·1, the significance of these is that they give "steady states" distinguished by the constancy of some observable characteristic. Primarily therefore the dynamical equations apply to the double space vectors T_0 of a particle and comparison fluid, since the particle alone has no observable characteristics. But in practice we substitute the simple space vectors T, T' which express (unobservable) relations of the two systems to a geometrical frame. Whether these will satisfy the equations, i.e. whether T and T' will be parallelly

† For brevity, the system whose strain vector is S will be called "the system S".

displaced when the double vector T_0 containing the recognisable characteristic is parallelly displaced, depends on their relation to T_0. If $T_0 = TT'$, parallel displacement of T and T' involves parallel displacement of T_0. But this is not in general true if $T_0 d\omega_0 = TT' d\omega d\omega'$, since volume elements are not transformed by parallel displacement in the same way as space vectors. The dynamical equations currently adopted therefore impose an additional condition $TT' = T_0$, or in the present application $TT' = \{T_0\}$. The corresponding strain vectors therefore satisfy $SS' = \{S_0\}$.

Equation (12·13) results from a comparison of two conditions,

$$\{S_0\} d\omega_0 = kS d\omega \,.\, S' d\omega', \quad \{S_0\} = SS'. \tag{12·14}$$

The first is imposed by the law of multiplication of probabilities, and the second by the dynamical equations.

We can now define more narrowly the circumstances in which (12·13) has to be satisfied. Since it represents an assumption in the dynamical equations, it applies to displacement in the dynamical coordinates. Further, since the dynamical equations are differential equations of the second order (§ 9·2), we have to retain squares of the coordinates in the investigation. Local orthogonal coordinates in phase space are insufficient; and it will be necessary to use a more extended system such as stereographic coordinates.

Before applying (12·13) we must set forth certain general considerations which arise in combining two systems. For purposes of exposition it is more convenient to consider two elementary particles S, S'; but S' will ultimately be replaced by a neutral particle.

12·2. Combination of Two Systems.

As an approach to the more general theory, let us first consider the combination of two particles S, S' into a single system S_0 according to the elementary wave mechanics of de Broglie and Schrödinger, in which each particle has only one phase variable. Let θ, θ' be the phase variables, so that the wave functions of S and S' are

$$\psi = e^{\frac{1}{2}i\theta} \psi_0, \quad \psi' = e^{\frac{1}{2}i\theta'} \psi_0'. \tag{12·211}$$

These are combined by multiplication, so that, setting $\Psi_0 = \psi_0 \psi_0'$, the wave function of the double system is

$$\Psi = e^{\frac{1}{2}i(\theta+\theta')} \Psi_0. \tag{12·212}$$

As in § 7·4, we introduce linear variables by setting

$$ds = R d\theta, \quad ds' = R' d\theta'. \tag{12·213}$$

Then, writing

$$m = 1/2R, \quad m' = 1/2R', \tag{12·214}$$

we have

$$\Psi = e^{i(ms+m's')} \Psi_0. \tag{12·215}$$

In this form the phase angle $ms + m's'$ can be interpreted as action, the action of a system being $\Sigma \int m\, ds$.

We have described Ψ as the wave function of a *double system*. That is not quite the same thing as the wave function of a *combined system*. To obtain the wave function of the combined system, the time coordinates t, t' of the two particles must be equated. Consider, for example, the sun-earth system. That does not comprise a combination of the earth today with the sun a week ago; no reference to the "orbit" of such a combination will be found in astronomical textbooks. The essence of the process of "combining" is the substitution of a single time coordinate for the whole system instead of independent time coordinates for its separate parts (§ 7·8).

We are not arguing that there would be anything illegitimate in the conception of S and S' as a double system, i.e. with two independent time variables. Our point is that there are two possible conceptions which must be carefully distinguished, and that the usual description of a composite system refers to the conception of it as a combined system, i.e. with a single time variable. When we speak of the orbit of a planet or the quantum state of an atom we are referring to the combined system, not the double system.

When s and s' are expressed as functions of the coordinates, the wave function Ψ of the double system is a function of eight coordinates $x, y, z, t, x', y', z', t'$. The wave function of the combined system S_0 is $(\Psi)_{t'-t}$, a function of seven variables. The mass or energy of S_0 is the value of the operator $-i\partial/\partial t$, which by (12·215) is

$$m\,ds/dt + m'\,(ds'/dt')_{t'-t}.$$

This is the sum of the masses of S and S' at a simultaneous instant, the factors ds/dt, ds'/dt' being the FitzGerald factors representing change of mass with velocity.

In the general theory the simple algebraic phases θ, θ' are replaced by space-like matrices with ten components (§ 7·3) which, for a range of configurations small enough to be referred to local orthogonal coordinates, may be denoted by

$$\left.\begin{aligned}\Theta &= \Sigma_s E_\mu \theta_\mu = \Sigma_s E_\mu x_\mu/R,\\ \Theta' &= \Sigma_s F_\mu \theta_\mu' = \Sigma_s F_\mu x_\mu'/R',\end{aligned}\right\} \tag{12·221}$$

and the wave functions (within the above small range) are

$$\psi = e^{\frac{i}{h}\Theta}\psi_0, \quad \psi' = e^{\frac{i}{h}\Theta'}\psi_0'. \tag{12·222}$$

For the wave function representing a combined system S_0 we have similarly

$$\Psi = e^{\frac{i}{h}\Phi}\Psi_0, \tag{12·231}$$

where

$$i\Phi = \Sigma_s E_\mu F_\nu \theta_{\mu\nu} = \Sigma_s E_\mu F_\nu x_{\mu\nu}/R_0. \tag{12·232}$$

In forming a combined system out of S and S' we equate their times t and t', and this common time is also the time t_0 of the combined system S_0. Since the time in phase space is represented by the algebraic coordinate, the identification $t = t' = t_0$ becomes, in the above notation,

$$x_{16} = x_{16}' = x_{16,16}. \tag{12·24}$$

We notice that, whereas in Schrödinger's theory the algebraic phase variable represents s, in the general theory it represents t. The difference is not surprising, since in the former theory one variable has to do the work of ten. In the matrix theory the strains of the system corresponding to displacements along x, y, z are provided for more precisely by separate non-algebraic phase variables, and therefore no longer enter into the algebraic phase. A possible misunderstanding may be caused by the fact that in matrix theory just as in Schrödinger's theory the wave function for plane waves is $\psi = e^{ims}\psi_0$; but here *ims* is the *eigenvalue of a matrix expression* representing the phases (ψ_0 having been specially chosen as an eigensymbol), and is not to be identified with the algebraic phase $\frac{1}{2}i\theta_{16}$.

It may be well to state again the reason why the time is represented by the algebraic coordinate in phase space. Progress in time corresponds to rotation (about the centre of curvature of space-time) in the plane associated with E_{45}. Hence energy conjugate to the time is the E_{45} component of the complete space vector which comprises the mechanical properties of the particle. The space vector is multiplied by iE_{45} to form the associated strain vector; energy is therefore the E_{16} component of the strain vector. The transformation $q = e^{\frac{1}{2}E_{16}\theta_{16}}$ applied to the strain vectors in phase space, which causes displacement in the algebraic coordinate, is therefore a displacement conjugate to the energy, i.e. displacement in time.

The phase space of S_0 contains 136 coordinates $\theta_{\mu\nu}$. It may be suggested that since the configurations of S and S' are each completely specified by ten coordinates, one of which they have in common, we shall require a phase space of 19 dimensions at most to specify the combined system S_0. But that is to begin at the wrong end of the problem. Our observational data relate to complex systems. We do not construct S_0 to represent data which have been ascertained about S and S'; we construct S and S' to represent data which have been ascertained about S_0. We have to take a system represented by a double wave vector in any combination of its 136 phases, all combinations being relativistically equivalent, and determine the conditions under which it can be dissected into two systems each with ten phases, or into a ten-phase system and a neutral comparison fluid with a single phase.

Initially the probability of S_0 is distributed uniformly over its 136-dimensional phase space. We replace it by two initial probability distributions uniform over the ten-dimensional phase spaces of S and S'. A great variety of configurations regarded as potentially distinguishable in the first representation must be classed as indistinguishable in the second representation. That is to say, there is an alteration of the basis of statistics. It will be found that this alteration has important consequences.

12·3. The Augmented Phase Space.

In combining S and S' we drop a time coordinate. Accordingly before dissecting S_0 we must insert an additional time coordinate and treat S_0 as having an infinitesimal uniform extension in this extra time. Let

$$t_0 = \tfrac{1}{2}(t+t'), \quad t_r = t - t', \tag{12·31}$$

so that

$$dt_0 dt_r = dt dt'. \tag{12·32}$$

Reducing to angular measure with the respective scale constants of the three phase spaces, (12·32) becomes

$$R_0^2 d\theta_0 d\theta_r = RR' d\theta_{16} d\theta_{16}'. \tag{12·33}$$

By this augmentation the phase space of S_0 contains two algebraic phases θ_0 $(=\theta_{16,16})$ and θ_r. Owing to their commutative property we can treat the probability distribution in the algebraic phases independently of the distribution in the other phases (cf. § 7·7). For example, in ten-dimensional phase space the volume element in local orthogonal coordinates $d\omega = d\theta_1 d\theta_2 \ldots$ contains an amount of probability $d\omega/\Omega$, where Ω is the whole volume of phase space which we have shown to be finite (§ 7·4). Writing $d\omega_c$ and Ω_c for the corresponding volumes without the algebraic dimension, we have

$$d\omega = d\theta_{16} d\omega_c, \quad \Omega = 2\pi\Omega_c, \tag{12·34}$$

and the probability $d\omega/\Omega$ is the product of the independent probabilities $d\theta_{16}/2\pi$ and $d\omega_c/\Omega_c$ that θ_{16} is in the range $d\theta_{16}$ and that the other coordinates are in the range $d\omega_c$.

Hence the initial probability associated with the range $d\theta_{16} d\theta_{16}'$ is $(d\theta_{16}/2\pi)(d\theta_{16}'/2\pi)$ and the initial probability associated with the range $d\theta_0 d\theta_r$ is $(d\theta_0/2\pi)(d\theta_r/2\pi)$. These must be equal, since $d\theta_{16} d\theta_{16}'$ and $d\theta_0 d\theta_r$ represent the same range described in different ways. Hence by (12·33)

$$R_0^2 = RR'. \tag{12·35}$$

The new variable θ_r must, like the other phases, correspond to a circular (not hyperbolic) transformation, since it is essential that phase space shall be closed (§ 7·3). We cannot begin the process of dissection of S_0 unless this condition is satisfied. It is, however, desirable to examine an argument which seems to suggest that t_r should be represented by a hyperbolic transformation.

The definition $\theta_r = t_r/R_0$ refers to local orthogonal coordinates, and for extended coordinate systems the relation of t_r and θ_r becomes non-linear. Adopting stereographic coordinates and considering variations of t_r only, we have, by (7·58),

$$d\theta_r = (1 + t_r^2/4R_0^2)^{-1} dt_r/R_0,$$

so that $\theta_r = 2\tan^{-1}(t_r/2R_0)$. Thus, as θ_r increases, t_r increases without limit,

provided that θ_r is real. If, however, θ_r is an imaginary quantity iu_r, we replace t_r by it_r and obtain

$$du_r = (1 - t_r^2/4R_0^2)\, dt_r/R_0,$$

so that $u_r = 2\tanh^{-1}(t_r/2R_0)$. Then as $u_r \to \infty$, $t_r \to 2R_0$. Accordingly $t - t'$ is not greater than $2R_0$. It suggests itself that $2R_0$ should be identified with the distance between S and S', and therefore with the light-time (the velocity of light being unity). The result $t - t' \not> 2R_0$ then means that two systems cannot be combined at instants such that one is in the absolute future of the other.

The variable u_r is found to be important in another problem which arises in the combination of two particles (§ 15·5), but here it is irrelevant. The radius R_0 of phase space is a scale constant applicable to all configurations represented in phase space. If we equate $2R_0$ to the distance between the particles, the phase space can contain only those configurations for which the distance is a fixed constant. That is not the problem here treated.

12·4. The Fundamental Quadratic Equation.

Writing $d\omega = \sqrt{-g}.d\tau$ as in (7·422), the relation (12·13) becomes

$$\sqrt{-g_0}.d\tau_0 = k\sqrt{-g}.\sqrt{-g'}.d\tau d\tau'. \tag{12·41}$$

The formula now applies to any system of coordinates.

We adopt stereographic coordinates. It has been shown in § 7·5 that when a transformation (not necessarily infinitesimal, but not involving anti-perpendicular components) is applied, each point of phase space in the infinitesimal neighbourhood of the origin receives the same increment of its stereographic coordinates. That is to say, when a volume element is displaced in phase space $d\tau$ remains constant. Hence for such displacements in the three phase spaces we must have

$$\sqrt{-g_0} = C\sqrt{-g}.\sqrt{-g'}, \tag{12·42}$$

where C is a constant. By (7·59)

$$\sqrt{-g} = R^{-n}(1 + r^2/4R^2)^{-n},$$

where r^2 is the square of the length of the displacement, and n is the number of dimensions of the phase space. Inserting this in (12·42), we have

$$(1 + r_0^2/4R_0^2)^{-n_0} = (1 + r^2/4R^2)^{-n}(1 + r'^2/4R'^2)^{-n'} \tag{12·43}$$

together with $R_0^{-n_0} = CR^{-n}R'^{-n'}$.

Consider a displacement in time, so that r, r', r_0 are x_{16}, x_{16}', $x_{16,16}$, and are all equal by (12·24). Then, expanding (12·43) and equating coefficients of r^2, we have

$$n_0/R_0^2 = n/R^2 + n'/R'^2. \tag{12·44}$$

By (12·35) and (12·214), $R_0^2 = RR'$, and $R/R' = m'/m$. Hence

$$nm^2 - n_0 mm' + n'm'^2 = 0. \tag{12·45}$$

For an elementary particle and a neutral comparison fluid, the dimensions of the phase spaces are

$$n=10, \quad n_0=136, \quad n'=1. \tag{12·46}$$

(The strain vector of a neutral particle, being algebraic, has only one phase variable.) Hence we obtain the fundamental quadratic equation

$$10m^2-136mm'+m'^2=0. \tag{12·47}$$

Its two solutions give two possible masses m for an elementary charged particle in terms of the mass m' of a neutral particle. These accordingly will be the masses of the proton and electron. The ratio of the two roots is 1847·6.

This is the ratio of the masses of the proton and electron, when mass or energy is defined as in quantum theory by the operator $(-ih/2\pi)\partial/\partial t$, or equivalently by $E=h\nu$. We shall find in § 15·9, that this deviates somewhat from the classical definition of mass and energy; and that, adopting the classical definition, the theoretical mass-ratio is 1834·1. Since all our formulae imply the quantum definition, we must adhere to the value 1847·6 in the developments which follow.

12·5. Notes on the Solution.

One or two steps in the foregoing proof require fuller discussion.

Should we take $n_0=136$ (original phase space) or 137 (augmented phase space)? Equation (12·42) was obtained by considering parallel displacement of a volume element $d\tau$ conformably with the dynamical equations. The angular element $d\theta_r$ is not a range dynamically transferred from the origin by parallel displacement; it is a constant infinitesimal thickness assigned to phase space, and hence equal for all volume elements considered. The value $n_0=136$ is therefore correct.

Is it legitimate to satisfy (12·43) as far as r^2, leaving a discrepancy in the higher powers of r? Equation (12·43) is not valid beyond r^2; its exact form would be

$$(1+t_0^2/4R_0^2)^{-n_0}=(1+t^2/4R^2)^{-n}(1+t'^2/4R'^2)^{-n'}, \tag{12·51}$$

where t_0, t, t' are corresponding times in the three systems *measured in stereographic coordinates*. The times are equal in some system of reckoning, but there is no reason to suppose that that reckoning is stereographic. In fact for large values of t the stereographic reckoning becomes absurd. The substitution $t_0=t=t'=r$ is only valid to the first order of small quantities, and therefore (12·43) is only valid to the second order. To the first order (corresponding to natural coordinates) the reckoning of t is fixed by the scale constant R, and there is no ambiguity as to what is meant by the three systems being simultaneous. In short, having found m/m' from the terms in t^2, the higher powers merely tell us how we must define simultaneity of the three systems for large values of t.

By obtaining a formula correct as far as r^2, we obtain exact values of the second derivatives at the origin; so that we achieve our aim of providing for the second order differential equations which express the laws of physics. Results obtained in a special coordinate system, such as stereographic coordinates, are of no great interest in themselves. We use them to extract the invariants, in particular the curvature invariants which correspond to mass and density. For this purpose it is only necessary to expand as far as r^2.

We come to a more difficult point. The displacement which we have considered is along the algebraic coordinate; but the algebraic coordinate is very much aloof from the others; and although we have used the same stereographic projection for it as for the others, there is no obvious need to do so. It is not possible in an extended region to use natural measure for all coordinates; but it is possible to use natural measure for the algebraic coordinate and stereographic measure for the rest. Geometrically the distinction is that the algebraic dimension has cylindrical curvature, whereas the others by their non-commutative relations determine a spherical curvature. Spherical curvature is an invariant; and on the assumption that R, R', R_0 are radii of spherical curvature, the result (12·44) is independent of the special system of coordinates used in obtaining it. Unfortunately, in the equation as derived, they are radii of cylindrical curvature. To validate the proof we must show that there is some condition (not yet mentioned) which requires the algebraic coordinate to be represented uniformly with the others.

The fact is that we have not explicitly introduced the condition that the particle whose mass we are seeking is a pure elementary particle, having therefore a singular strain vector. As shown in § 7·9 such a particle will have, associated with it, a singular line in phase space given by $d\Theta_s = (-E_1 + E_{16})\, d\phi$. This provides an absolute scale comparison of lengths in the E_1 and the E_{16} directions—just as in four-dimensional space-time a singular line (constituting a light track) provides an absolute scale comparison between intervals of space and time.

To go back to the first principles of measurement—magnitudes in different directions are compared on the basis that "equivalent" magnitudes (which can be transformed into one another by relativity rotations) are equal magnitudes. This enables us to compare magnitudes in perpendicular directions immediately, and magnitudes in antiperpendicular directions indirectly, using for the latter a direction perpendicular to both as intermediary. But this does not apply to the algebraic direction which has no direction perpendicular to it. The measurement of proper mass, proper energy, proper time, in terms of the standards used for measuring momenta and lengths in other directions, must depend on a different principle of comparison. This is supplied by the formula (6·64) for the strain vector of a pure

particle. Its momenta in four antiperpendicular directions E_1, E_{23}, E_{45}, E_{16} are equal. This enables us to compare the momenta in the E_{16} direction with the momenta in other directions. If, as suggested above, a different scale constant R were employed for the E_{16} direction, a stream vector which is pure in angular measure would not be pure in linear measure; and our whole treatment of the analysis of matter into pure wave functions representing elementary particles would have to be revised.

We may put the result in another way. An elementary particle possesses spin. We can measure time by the spin coordinate of the particle—as time is in fact measured by the spin coordinate of the earth. Thus x_1, x_{11} can be substituted for x_{16}, $x_{16,16}$ in our previous deduction; and the difficulties arising from the peculiar character of the algebraic coordinate are then evaded.† This brings out the fact that the mass which we have determined is that of an elementary particle. A distinctive feature of the elementary particle is that it has four dynamical coordinates all of which measure the time; but two of them are not shown in phase space.

12·6. Energy Invariants.

The primary operation of wave mechanics is multiplication; and in the combination of two systems the product of the masses $m_1 m_2$ is more fundamental than the sum $m_1 + m_2$. This was illustrated in the transformation to relative coordinates, in which the original particles are replaced by external m and internal particles whose masses M, μ satisfy $M\mu = m_1 m_2$.

We shall call $m_1 m_2$ the *mutual pressure invariant* of the two particles. Correspondingly m_1^2 is called a *self-pressure invariant.*

Let us provisionally interpret the term "pressure" literally. In statistical mechanics the pressure is the energy associated with one degree of freedom.‡ If there is equipartition between n degrees of freedom, the energy is $nm_1 m_2$. We shall call $nm_1 m_2$ the *mutual energy invariant.* The *self-energy invariant* is defined similarly.

The fundamental equation (12·45)

$$n_0 mm' = nm^2 + n'm'^2$$

expresses that the mutual energy invariant of the system S_0 is the sum of the self-energy invariants of the two parts composing it. We thus get a kind of physical picture of the significance of the fundamental quadratic. In particular the equation§

$$136mm_0 = 10m^2 + m_0^2 \qquad (12{\cdot}61)$$

† The algebraic coordinate was employed because we knew what were "corresponding" displacements of the three systems in that coordinate but not in any other coordinate. But the above argument shows that, besides (12·24), we have $x_1 = x_{11}$.

‡ We have in mind waves in which the energy is half kinetic and half potential. If kinetic energy alone is present a factor $\frac{1}{2}$ is introduced.

§ We shall in future denote the mass of the neutral comparison particle by m_0 instead of m'—since m' seems a rather unsuitable notation for an important constant of nature.

means that the introduction of the elementary particle m can be regarded either as *replacing* a neutral unspecified particle m_0 by a double system with 136-dimensional phase space or as *adding* a simple system with ten-dimensional phase space. Both interpretations give the same total energy invariant. From this point of view the number of dimensions of phase space is involved, because in the steady state represented by the initial probability distribution, each dimension shares in the equipartition of energy.

Let us extend this to a static system formed by two elementary particles with a neutral comparison fluid. We know that two protons or two electrons cannot form a static system; and we may therefore anticipate that the two particles will be a proton and electron. In short, our system is a hydrogen atom (or possibly a neutron).

We resolve the hydrogen atom as in § 10·9 into an external and internal particle, described by external and internal wave functions, with masses

$$M = m_p + m_e, \quad \mu = m_p m_e/(m_p + m_e). \tag{12·62}$$

Since m_p, m_e are the roots of $10m^2 - 136mm_0 + m_0^2 = 0$, we have

$$M = (136/10)\, m_0, \quad \mu = (1/136)\, m_0.$$

Writing these in the form

$$136 M m_0 = 10 M^2, \quad 136 \mu m_0 = m_0^2, \tag{12·63}$$

we see that for an external particle the mutual energy invariant of the particle and its comparison fluid is equal to the self-energy invariant of the particle; but for the internal particle the mutual energy invariant is eq to the self-energy invariant of the comparison fluid. The two particles have therefore a very different type of relationship to the comparison fluid.

We can extend the foregoing definition of pressure and energy invariants to conditions in which the energy is represented by a symbolic operator W. The self-energy invariant of a particle in phase space of n dimensions is then nW^2. If the wave functions representing steady states satisfy Dirac's equation $W\psi = m\psi$, it follows that $(W^2/m)\psi = m\psi$. Thus we have two operational forms of the energy m, namely the linear hamiltonian W and the second order hamiltonian W^2/m. For the external particle

$$\frac{W^2}{M} = \frac{10W^2}{136m_0} = \frac{1}{\gamma_e} nW^2 \quad \text{(external particle)}, \tag{12·641}$$

where γ_e is the natural constant $136m_0$, and n is the number of dimensions of the phase space of the external particle. For the internal particle

$$\frac{W^2}{\mu} = \frac{136W^2}{m_0} = \frac{1}{\gamma_i} nW^2 \quad \text{(internal particle)}, \tag{12·642}$$

where

$$\gamma_i = \gamma_e/136^2$$

and n is unity. Thus γ_i, γ_e are the factors which convert energy invariants into energies on the ordinary scale.

For an electron or proton, the fundamental quadratic can be written as

$$m = 10m^2/\gamma_e + m_0^2/\gamma_e,$$

so that the operational form of the energy is

$$\frac{1}{\gamma_e}(nW^2 + m_0^2) \quad \text{(elementary particle)}. \tag{12·643}$$

It is interesting to compare (12·643) and (12·641).

I think that for theoretical purposes the energy invariant is to be preferred to the energy; that is to say, we should replace the quadratic energy operator W^2/m by nW^2/γ. In the following chapters I have not used the energy invariants as much as I might have done, fearing that it would be a stumbling-block to the reader. When the form W^2/m is used, it should be realised that m is a combination of the universal constant m_0 with a geometrical factor of the problem, the latter introducing the number of degrees of freedom. It is also particularly important to notice that m is not necessarily the rest energy; for internal wave functions (which are those most concerned in practical problems) the rest energy is zero (10·982).

12·7. The Association of Mass and Charge.

We have found that an elementary particle must have one of two masses m_p, m_e. We shall now show that in static conditions the particles of mass m_p are of one sign (arbitrarily called positive), and the particles of mass m_e are of the other sign; also that the positive and negative charges are equal.

To test the sign of the charge we introduce a field of uniform electrostatic potential κ_4. If we can show that the effect of the field is to make the mass m^* of an elementary particle at rest satisfy the modified equation

$$10(m^*)^2 - 136m^* m_0 + \sigma^2 m_0^2 = 0, \tag{12·71}$$

where σ is a function of κ_4, the required conclusion follows. For then

$$m_p{}^* + m_e{}^* = \tfrac{136}{10} m_0 = m_p + m_e, \tag{12·72}$$

so that we may write $\quad m_p{}^* = m_p + \epsilon, \quad m_e{}^* = m_e - \epsilon, \tag{12·73}$

showing that the part of the mass or energy due to the existence of the field is equal and opposite for the two types of particle. Our problem therefore reduces to showing that κ_4 affects only the last term of the quadratic.

By § 8·8 the constant potential κ_4 can be removed by applying the gauge transformation $\psi' = e^{i\kappa_4 x_4}\psi$ to wave vectors of index 1. The strain vector of index 2 generating phase space then undergoes the transformation

$S' = e^{2i\kappa_4 x_4} S = e^{-2\kappa_4 t} S$. S already contains the time factor $e^{i\theta_{10}} = e^{it/R}$; so that the time factor of S' is $e^{it'/R}$, where

$$t' = t(1 + 2i\kappa_4 R) = t(1 - 2\kappa_0 R). \tag{12·74}$$

Thus the field-disturbed environment of the object system is converted into the standard environment (neutral space-time) by antedating our description of it.† In analysing (§ 12·4) the double wave functions, which contain the observable relations, into a simple wave function of the particle and a one-phase wave function of the comparison fluid, we gave to all three functions the same time t; if, however, we give the one-phase wave function a time t', it will equally represent a neutral comparison fluid at time t' or the actual field-disturbed environment at time t. With this modification (12·43) becomes

$$(1 + t^2/4R_0^2)^{-n_0} = (1 + t^2/4R^2)^{-n} (1 + t'^2/4R'^2)^{-n'} \tag{12·75}$$

and gives the condition for analysis of the double wave functions, containing the observable relations of the particle to its field-disturbed environment at time t, into a simple wave function of the particle and a one-phase wave function of its field-disturbed environment at the same time t. The change only affects the last term of the quadratic equation, which becomes

$$nm^2 - n_0 mm' + (t'/t)^2 m'^2 = 0. \tag{12·76}$$

This is the result we required.

We do not apply the gauge transformation to the double wave function, because that contains the observable relations, and it is understood that the energy of the particle is to be determined from observation precisely as if the actual environment were the standard environment. As we should ordinarily say, the change from m to m^* is an apparent change of mass, due to the difference between the comparison fluid actually present and available for our observations and the ideal comparison fluid referred to in the definition of m.

The value of σ in (12·71) is $(1 - 2\kappa_0 R')$; or, since $1/2R' = m_0$ (formerly called m'),

$$\sigma = 1 - \kappa_0/m_0. \tag{12·771}$$

Using (12·73), we derive the relation

$$\left(1 + \frac{\epsilon}{m_p}\right)\left(1 - \frac{\epsilon}{m_e}\right) = \left(1 - \frac{\kappa_0}{m_0}\right)^2. \tag{12·772}$$

Thus σ is real for a physically real electrostatic potential, and (for electromagnetic fields which are not too extravagantly great) ϵ is also real. By a happy accident, positive electric potential corresponds to positive κ_0.

The numerical value of σ does not much concern us; but if any application

† The reader may perhaps be inclined to object that, if (as usual) the comparison fluid is at rest, antedating makes no difference. But there is a phase angle which changes with the time; and we see above that it is this phase angle which exhibits the antedating.

is made of (12·771), we should notice that κ_0 is not there measured in the same units as in the ordinary wave equation since it refers to a gauge transformation of the comparison fluid.

It is interesting to consider the physical basis of the dissymmetry of positive and negative charges. Formally they are correlated to the two square roots of -1. This can only give rise to observable dissymmetry (such as a difference of mass) if, elsewhere in the theory, the two square roots have been allotted unsymmetrical rôles in the description of phenomena. In the expression for a physically real interval $E_1x_1+E_2x_2+E_3x_3+E_4it$ we have to select one of the square roots of -1, which we shall denote by i_f, such that E_4i_ft represents displacement towards the future when t is positive. Then any other square root of -1 occurring in our formulae is distinguished as i_f or $-i_f$.

The irreversibility of time is manifested in three ways: (*a*) in consciousness, (*b*) in the laws of entropy, (*c*) in the cosmic expansion. Consciousness must be regarded as the ultimate source of the irreversibility, at any rate from the point of view of physical theory. Owing to the curious fact that our minds are acquainted by sensory mechanism with the past but not the future, observational knowledge has the form of an integral over past time up to the present moment t. Since probability is relative to knowledge, a formulation of the universe in terms of probability distributions is exceedingly unsymmetrical with respect to past and future time; and this irreversibility is shown in the manifestations (*b*) and (*c*). It is also shown in wave mechanics in a more elementary way by the concentrated wave packets, which are formed discontinuously by our observations and diffuse continuously as t increases towards the future.

For the distinction of protons and electrons the most relevant manifestation is the cosmic expansion. I do not mean the actual "expansion of the universe", which might perhaps have been reversed by altering the initial conditions. But a region of space with de Sitter metric necessarily expands relatively to our standards of measurement. We generally treat a small region of space isolated from the rest by artificial boundary conditions; but the dilemma, first pointed out by de Sitter, always appears—either we must take it to be a portion of static spherical space (an Einstein universe), in which case Lorentz transformations are inapplicable; or we must take it to be a portion of de Sitter space-time, in which case geodesics that are initially parallel proceed to diverge. Since we have introduced space-time by Lorentz transformations, our method follows the latter alternative. Correspondingly we have an expanding comparison fluid, which is unsymmetrically related to past and future time. In particular, elementary particles symmetrically related to the expanding comparison fluid are not symmetrically related to past and future time.

12·8. The Stern-Gerlach Effect.

We shall now calculate the energy of a hydrogen atom in an external macroscopic electromagnetic field. Representing the atom by external and internal wave functions, we have to calculate the effect of the field on these separately. An essential point in the investigation is that the external wave functions are continuous and the internal wave functions are discrete.

We use the method of gauge transformations explained in § 8·8. Let us introduce a field of electromagnetic potential by changing the local unit of length, so that the scale constants R, R' of the phase spaces of a particle and its comparison fluid become in the new measure $\beta^{-1}R$, $\beta^{-1}R'$, where β is a unitary complex factor. The coefficient $\sqrt{-g}$ in the reduction from coordinate volume to natural volume will be changed to $\beta^{10}\sqrt{-g}$ for the ten-dimensional phase space and $\beta\sqrt{-g}$ for the one-dimensional phase space. We do not recognise complex measures of volume; accordingly the factors are transferred to the strain vectors which become $\beta^{10}S$, $\beta S'$. The transfer means that the change of probability distribution due to the electromagnetic field is incorporated in the modifying factor instead of in the initial probability distribution.

The masses of the particles are such as to validate the usual assumption that the volume elements can be ignored in the dynamical equations; that is to say, S and S' can be treated as discrete strain vectors in the dynamical equations. But they do not behave as discrete wave vectors for gauge transformations, and consequently the electromagnetic terms in the dynamical equations are not the same. Let us now consider the gauge transformation for a discrete strain vector.

A discrete wave function exists only for discrete values of its parameters; but it is a continuous function of coordinates, so that it is affected by the gauge transformation of the volume element of coordinates. By the theory of Chapter VIII a discrete wave function has four dynamical coordinates; whence it follows that the transformation of the strain vector is $S \to \beta^4 S$. We obtain the same result more directly from the ordinary expression

$$\psi\phi^*\sqrt{-g}\,dx_1dx_2dx_3 \qquad (12{\cdot}81)$$

for the probability in an element $dx_1dx_2dx_3$ of three-dimensional space. Here $\sqrt{-g}$ is the four-dimensional factor, defined in ordinary tensor calculus, which varies as β^4.† The factor β^4 is therefore attached to the discrete strain vector function $\psi\phi^*$.

We have therefore the following result. Although the external wave function is continuous, it is permissible to treat it as discrete for the ordinary purposes of wave mechanics, the mass μ assigned to it having been expressly chosen so as to validate this practice. But if a gauge transformation is

† *Mathematical Theory of Relativity*, equation (85·44).

applied, the strain vector will be changed in the ratio β^{10} whereas the change for a discrete wave function would have been in the ratio β^4. The wave functions ψ, ϕ are therefore changed in the ratio β^5, as compared with β^2.

Applying the theory of § 8·8 with $\lambda = \beta^2, \beta^5$, if the potentials in the wave equation for a discrete wave function are κ_μ, those in the equation for the external wave function are $\frac{5}{2}\kappa_\mu$. Thus for the same electromagnetic field we have

$$p_\mu = -i\frac{\partial}{\partial x_\mu} + \kappa_\mu \quad \text{(discrete wave functions)}, \tag{12·821}$$

$$p_\mu = -i\frac{\partial}{\partial x_\mu} + \tfrac{5}{2}\kappa_\mu \quad \text{(continuous wave functions)}. \tag{12·822}$$

The hamiltonian of the external particle is therefore

$$W = \sum_1^4 E_\mu(-i\partial/\partial x_\mu + \kappa_\mu') \quad \kappa_\mu' = \tfrac{5}{2}\kappa_\mu, \tag{12·831}$$

which gives the second order hamiltonian

$$\frac{W^2}{M} = \frac{1}{M}\Sigma\left\{\frac{\partial^2}{\partial x_\mu^2} - \kappa_\mu'^2 - iE_{\mu\nu}\left(\frac{\partial \kappa_\nu'}{\partial x_\mu} - \frac{\partial \kappa_\mu'}{\partial x_\nu}\right)\right\}. \tag{12·832}$$

If the field consists of a magnetic force $eH_1 = \partial\kappa_2/\partial x_3 - \partial\kappa_3/\partial x_2$ in the x_1 direction, the last term becomes $(\frac{5}{2}ieH_1/M)E_{23}$. To the classical approximation this term is twice the mutual energy of the particle and electromagnetic field. The mutual energy is thus

$$\frac{5}{2}\cdot\frac{eH_1}{M}\cdot\tfrac{1}{2}iE_{23}. \tag{12·84}$$

The factor $\frac{1}{2}iE_{23}$ is the spin momentum of the external particle in the plane normal to H_1 and has eigenvalues $\pm\frac{1}{2}$. The factor $\frac{5}{2}e/M$, or in the usual units $\frac{5}{2}e/Mc$, is the effective magnetic moment per unit angular momentum.

The corresponding result for the internal particle is

$$\frac{eH_1}{\mu}\cdot Z, \tag{12·85}$$

where the operator Z includes, besides the spin momentum, half the orbital momentum.† The factor $\frac{5}{2}$ is now omitted because the internal wave functions are discrete. But it is necessary to consider whether they are the standard discrete wave functions in *four* dimensions referred to in (12·821), or whether the fact that internal space is three-dimensional will not give $\kappa_\mu' = \frac{3}{4}\kappa_\mu$, and an energy

$$\frac{3}{4}\cdot\frac{eH_1}{\mu}\cdot Z. \tag{12·86}$$

I think that for the theory of the atom, so far as it is developed in Chapter IX, (12·86) is the correct result, the factor $\frac{3}{4}$ being essentially the same as that which will occur later in (14·153). It is true that the internal wave function has the standard number of dynamical coordinates, and appears

† The factor $\frac{1}{2}$ is the well-known magnetic anomaly.

in §§ 9·2, 9·3 to be a function of t as well as of the space coordinates. But although t is used in Chapter IX in order to agree with the usual notation, t is not the time; it will be shown in Chapter XV that it is an interchange coordinate. Now the interchange coordinate is by its nature gauge-invariant. Thus for *internal* space, $\sqrt{-g}$ varies as β^3 instead of β^4, and the factor $\frac{3}{4}$ is required.

But it is well known that, when a magnetic field is applied to an atom, we have to distinguish between weak and strong fields. A strong field cannot be treated as a perturbation of the normal states, but involves a re-analysis into elementary states. The particle is, as it were, torn in its allegiance between the planes of simultaneity determined by the external momentum vector and those determined by H_1 (with time direction so chosen that the field is purely magnetic). This effect of strong fields is generally described as an "uncoupling" of spins. The magnetically determined planes of simultaneity introduce a genuine time coordinate into the internal state additional to the proper time s correlated to the external momentum vector; the original t (interchange coordinate) is then relegated to the rôle of argument for small perturbations, and the states are re-analysed with respect to the genuine time coordinate. It appears therefore that (12·85) is right for strong fields; but it will be replaced by (12·86) in weak fields, where a gauge-invariant interchange coordinate takes the place of the genuine time coordinate.

It is to be understood that in our equations H_1 is not necessarily a measure of the field in absolute units; it is used to connect (12·84) alternatively (but not simultaneously) with (12·85) or (12·86). Actually in passing from (12·85) to (12·86) we should change the unit of energy in proportion to the number of degrees of freedom, so that they represent the same absolute energy.

By suitable arrangements a stream of particles, projected in a strong inhomogeneous magnetic field, can be made to divide itself according to the different combinations of eigenvalues of (12·84) and (12·85). From the measured deflections the ratio of the factors $\frac{5}{2}e/Mc$ and $e/\mu c$ can be determined. The generally accepted results give the numerical coefficient $\frac{5}{2}$, agreeing with our theory. Since M and μ are very nearly equal to m_p and m_e, the magnetic energies are generally attributed to the proton and electron, respectively. But it is clear from the foregoing investigation that they properly belong to the external and internal motions, respectively.

The result for weak fields also appears to be confirmed by experiment (Rabi, Kellogg and Zacharias, *Physical Review*, **46**, 157 (1934)). In this case the ratio is $\frac{5}{2}e/Mc : \frac{3}{4}e/\mu c$, so that the numerical coefficient is $\frac{10}{3}$. The experimental result is stated to be $3{\cdot}25 \pm 0{\cdot}3$.

Whilst the result for strong fields is a simple consequence of the theory, the theory for weak fields may perhaps require a closer scrutiny than I have been able to give.

CHAPTER XIII

STANDING WAVES

13·1. Scalar Wave Functions.

Vector wave functions have superseded the scalar wave functions of Schrödinger in the problems of atomic physics; but there are certain applications of wave mechanics in which scalar wave functions are still needed. These will now occupy us for two chapters. We shall develop the theory of scalar wave functions independently, and afterwards show how the vector wave functions hitherto treated are connected to them.

We have seen that a neutral particle (neutral both as regards charge and spin) has an algebraic strain vector (§ 6·6); in this case the strain vector S becomes a strain scalar. Following the general procedure in wave mechanics, we represent it as the product of two scalar wave functions $\psi\phi$. We may expect that scalar wave functions will suffice to represent any macroscopic distribution which is neutral as regards charge and spin—in fact, such distributions as are fully specified by an energy tensor $T_{\mu\nu}$. If $T_{\mu\nu}$ is inadequate and the distribution is of a type represented by a general Riemann-Christoffel matrix (§ 11·4), vector wave functions will be required. The general theory of the representation of $T_{\mu\nu}$ by scalar wave functions is given in § 13·7. However, it is not of primary importance to ascertain the precise limits of the application of scalar wave functions; the main consideration is that they occur in certain problems of great importance in connecting relativity theory with quantum theory.

In macroscopic theory, energy-density, momentum-density and pressure are components of an energy tensor $T_{\alpha\beta}$. It is part of the definition of these quantities that they satisfy the law of conservation $(T^{\alpha\beta})_\beta = 0$; and by this property $T_{\alpha\beta}$ is identified with a geometrical tensor

$$-8\pi\kappa T_{\alpha\beta} = G_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}(G - 2\lambda), \qquad (13\cdot11)$$

which satisfies the law of conservation identically. We insert the constant κ (constant of gravitation) in order that we may be free to use the unit of mass, defined by $p_\mu = -i\partial/\partial x_\mu$, which simplifies the formulae of wave mechanics.

Here λ may be any constant, a change of λ being a change of the zero from which energy and pressure are reckoned. But the zero condition, $T_{\alpha\beta} = 0$, should not be taken to correspond to entire absence of matter and radiation (or in wave mechanics to zero probability of the presence of particles and photons). For if we adopt this as the zero condition *everywhere*, we are comparing the actual universe with a universe completely devoid of matter

and radiation and therefore without observable properties. We cannot adopt as a reference standard for physical measurements a system in which all such measurements become indeterminate. We shall later consider the convention adopted for fixing λ in current practice.

To make a transition to wave mechanics, $T_{\alpha\beta}$ must be interpreted as the expectation value of an operator $\mathbf{T}_{\alpha\beta}$. The only appropriate operation is covariant differentiation, and to satisfy tensor conditions we adopt

$$\mathbf{T}_{\alpha\beta} = -\frac{1}{m}\frac{\delta^2}{\delta x_\beta \delta x_\alpha}, \tag{13·12}$$

where δ stands for covariant differentiation, and m is an invariant. We shall be concerned only with scalar wave functions and natural coordinates; so that in our applications the covariant derivatives can be replaced by ordinary derivatives.

No hypothesis is implied in (13·12). It *defines* the wave functions which will be used for the description of macroscopic systems, viz. they are to be such as to give at each point an expectation value of $\mathbf{T}_{\alpha\beta}$ which agrees with the tensor defined in (13·11). The information contained in $T_{\alpha\beta}$ is thereby transferred to a pair of wave functions and made available for treatment by the methods of wave mechanics.†

Equation (13·12) is consistent with the usual definitions of momentum

$$\mathbf{p}_\alpha = -i\delta/\delta x_\alpha, \quad p^\alpha = m\,dx_\alpha/ds. \tag{13·131}$$

For, substituting in (13·12), we have

$$\mathbf{T}_{\alpha\beta} = \mathbf{p}_\beta \mathbf{p}_\alpha / m, \tag{13·132}$$

so that, when the momenta reduce to eigenvalues,

$$T^{\alpha\beta} = \frac{p^\alpha p^\beta}{m} = m\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds}, \tag{13·133}$$

which is a well-known form of the energy tensor of a system of particles when the right-hand side is summed in an appropriate invariant way.‡ Here, however, we regard (13·12) as the fundamental formula, not implying analysis into particles with eigenmomenta, but directly translating the curvature invariants of macroscopic theory into the operational forms of wave mechanics.

The coefficient m should not be prematurely identified with the proper mass of the system represented by the wave functions. As it is of the dimensions of mass, we may call it the mass-constant of the system. In most applications the units are adjusted so that, for an *elementary* wave system, m is the proper mass of the corresponding particle. For a composite wave

† In § 13·7 we shall follow the inverse procedure. The wave functions are defined directly, and the operational form $\mathbf{T}_{\alpha\beta}$ required to represent $T_{\alpha\beta}$ is deduced.

‡ *Mathematical Theory of Relativity*, equation (53·1).

system m can be interpreted as the proper mass of a single equivalent particle—which is by no means the same thing as the sum of the masses of the particles. For the purposes of general theory the division by m is rather inappropriate, and it would be better to write (13·12) as

$$m\mathbf{T}_{\alpha\beta} = -\delta^2/\delta x_\beta \delta x_\alpha, \tag{13·125}$$

defining a pressure invariant operator $m\mathbf{T}_{\alpha\beta}$ (cf. § 12·6).

The expectation value is formed in the usual way, namely,

$$T_{\alpha\beta} = \phi \mathbf{T}_{\alpha\beta} \psi. \tag{13·141}$$

Thus tensor conditions are satisfied if ϕ and ψ are invariants. For a small three-dimensional volume dV

$$T_{\alpha\beta} dV = \phi \mathbf{T}_{\alpha\beta} \psi dV. \tag{13·142}$$

If a component $\mathbf{T}_{\alpha\beta}$ reduces to an eigenvalue $t_{\alpha\beta}$, we have

$$T_{\alpha\beta} dV = t_{\alpha\beta} . \phi\psi dV. \tag{13·143}$$

This is conveniently regarded as expressing that the $\alpha\beta$-component of the energy in dV is due to a probability $\phi\psi dV$ that an entity with energy component $t_{\alpha\beta}$ is within the volume. We call this entity a *scalar particle*. It will be found later that an elementary scalar particle is equivalent to four elementary charged particles.

If the wave functions of the distribution are analysed into a set of elementary orthogonal wave functions, each of which is a discrete eigenfunction of certain components of $\mathbf{T}_{\alpha\beta}$ (in practice, the diagonal components) sufficient in number for the eigenfunction to be defined uniquely by their eigenvalues, we can introduce an *exclusion principle*. The exclusion principle states that in the whole domain of V there is not more than one scalar particle for each eigenfunction; so that the whole probability attached to any set of eigenvalues is not greater than 1. Using normalised eigenfunctions ψ_n, ϕ_n which satisfy

$$\int_V \phi_n \psi_n dV = 1, \tag{13·144}$$

ψ_n and ϕ_n will occur in the general wave functions $\Sigma a_n \psi_n$, $\Sigma a_n' \phi_n$ with coefficients not exceeding 1. Or, if $\psi_n' = a_n \psi_n$, $\phi_n' = a_n' \phi_n$,

$$\int_V \phi_n' \psi_n' dV \leqslant 1. \tag{13·145}$$

The foundation of this exclusion principle will be investigated in Chapter XVI. Meanwhile we accept the general idea of such a principle from current quantum theory.

The important point now arises that, since ψ_n, ϕ_n are invariants, the condition (13·145) is not invariant for Lorentz transformations or other relativity transformations which change dV. Thus, if it applies at all, it

applies to a special time axis. The particle interpretation is subject to the drawback, that by changing the time axis, the probability of the particle's existence in the domain V may become greater than unity! This has been urged as an objection to the use of scalar wave functions; and Schrödinger's development of them has been condemned as non-relativistic. Briefly the criticism is that, in order that the probability $\phi\psi dV$ of the particle being in a specified three-dimensional volume may be invariant for transformations of the coordinates, $\phi\psi$ must be a vector in the direction normal to dV. Therefore ϕ and ψ cannot be scalar quantities, but must be wave vectors whose product yields a space vector.

Much of Schrödinger's theory is non-relativistic; but in this instance the condemnation is not justified. The criticism overlooks that, whereas (13·142) is a tensor equation, invariant for all relativity transformations, (13·143) is definitely not invariant for *any* continuous transformation, since it is restricted by the non-invariant condition that ψ corresponds to the eigenvalue $t_{\alpha\beta}$. By hypothesis the eigenvalues defining ψ are discrete; otherwise no question of an exclusion principle arises. If we make an infinitesimal transformation of the coordinate system, the adjacent values of $t_{\alpha\beta}$ are impossible as eigenvalues, and ψ does not exist. It is a consequence of the discreteness of the eigenvalues that (13·143) can apply only to special orientations of the axes.

The physical reason is evident. Discrete eigenvalues arise from boundary conditions; and the components of the energy tensor which are quantised are those which have a special relation to the characteristics of the boundary. In particular, if the boundary is of constant form, it determines a space-time frame with respect to which it is at rest. The time component of the energy tensor thus becomes separated from the other components.

We have said that (for the present) we base the general idea of an exclusion principle on current quantum theory. As formulated for atomic systems it is amply confirmed by experiment. Undoubtedly an exclusion principle of some kind is obeyed by macroscopic distributions, and it is important to discover its precise formulation. There has not been the same opportunity for experimental test,† and formulae are current which have not been quantitatively verified. It is therefore necessary to examine the principle critically.

It is agreed that the unit wave functions (commonly regarded as representing individual particles) are eigenfunctions of a certain number of independent operators U_μ, and that the function is determined uniquely when the eigenvalues u_μ are specified. Also the eigenvalues are discrete,

† I think that the calculation, which we shall make later, of the constant of gravitation and the cosmical constant provides the first quantitative observational test. A qualitative test has been provided by the phenomenon of white dwarf stars.

although in some applications they may be very numerous and fall close together. It is therefore impossible to vary an eigenstate continuously; and if the distribution is given a small velocity, as in a Lorentz transformation, it must cease to be an eigenstate. It is therefore a necessary condition for an exclusion principle that it shall *not* satisfy the Lorentz transformation. Effectively this means that it applies to the relative or internal coordinates of a system—as is illustrated in the application to the hydrogen atom. For a neutral macroscopic distribution the choice of operators U_μ is very limited, since the only recognised characteristic of the distribution is $T_{\alpha\beta}$. In familiar applications ϕ and ψ are taken to be eigensymbols of the four diagonal components of $T_{\alpha\beta}$; so far as I can see, the only other possibility would be to employ derivatives (more especially curls) of $T_{\alpha\beta}$.

For the most part, what we have said about Lorentz transformations applies also to ordinary rotations. But when, owing to the symmetry of a system, it is impossible to distinguish its orientation with respect to its surroundings a degeneracy occurs; and the discrete set of eigenfunctions ϕ, ψ is defined by operators U_μ which include angular momenta.

There is no ground for supposing that an exclusion principle applies when the distribution is not in a steady state; and it seems evident that it does not apply. The energy concentrated in one elementary state of an organ pipe may (temporarily) be very much greater than $h\nu$ (§ 13·5). Consequently the principle cannot be extended from discrete to continuous wave functions, since the latter postulate an unbounded region in which equilibrium cannot be realised. When the eigenstates are very close together we may, for mathematical convenience, treat them as continuous, e.g. the states of a hydrogen atom just below ionisation; but this statistical continuity must be distinguished from the genuine continuity of the eigenstates which begins just above the ionisation level. If the ionised hydrogen atom is in a finite enclosure, the free electron can attain a steady probability distribution; but the eigenstates have again become discontinuous owing to the boundary conditions imposed by the enclosure.

The exclusion principle is often stated in the form that there is not more than one electron per unit cell ($\frac{1}{2}h^3$) of phase space. This is the most convenient form for statistical purposes. But since it no longer explicitly requires the wave functions to be discontinuous, many physicists have assumed that it applies to systems represented by progressive waves, which are not in statistical equilibrium. Any proposal to extend formulae relating to statistical equilibrium to non-equilibrium conditions requires justification; but no such justification has been attempted, and there is neither theoretical nor experimental ground for the assumption.

We add a comment on the association of particles with the unit wave functions. We are not attempting to analyse the macroscopic distribution

into the elementary particles, or even the neutral particles, treated in microscopic theory. Any particles which we now introduce are average particles, having composite individuality; and as such they have different properties from the elementary particles. It is permissible to regard an electron of composite individuality as an electron, in the same sense in which it is permissible to regard the Prime Minister as a human being—although he is subject to changes of size and appearance impossible in a human being as biologically defined, and might undergo displacement from Lossiemouth to Bewdley without traversing any intermediate region. An average particle is the particle which is fulfilling a particular rôle; and the conception arises when there is in the assemblage always one particle, or a probability of one particle, playing that rôle; but the identity of the particle is continually (and, for statistical purposes, continuously) changing. We shall find later that it is impossible even in microscopic theory to assign identity to particles at different times in an absolute way. But it is here unimportant whether the continuous identity which we assign to a microscopic particle is absolute or relative; adopting the microscopic particle as standard, the identity of the average particle associated with a macroscopic eigenfunction is composite and continuously changing. Naturally therefore it does not obey the same laws as a microscopic particle, in so far as the laws involve $\partial/\partial t$. Since $\partial/\partial t$ gives the mass and energy, the mass and hamiltonian are different. There are, however, well-known theorems concerning the density of distribution of eigensolutions, which normally secure that the number of particles found in different modes of dissection is invariant. We are therefore able to state the number of elementary particles corresponding to a given number of average particles.

13·2. The Box Problem.

Consider a rectangular block of matter (e.g. in the interior of a star) of dimensions $l_1 \times l_2 \times l_3$, and containing $\sigma l_1 l_2 l_3$ electrons. We wish to determine the *minimum* electron pressure P_0 as a function of the electron density σ.

We take axes such that the block as a whole is at rest, and we analyse the internal state of the block into standing waves in the relative coordinates ξ_1, ξ_2, ξ_3. This is done by resolving the scalar wave function (within the block) into Fourier components, so that $\psi = \Sigma \psi_{n_1,n_2,n_3}$, where

$$\left.\begin{aligned} \psi_{n_1,n_2,n_3} &= a_{n_1,n_2,n_3} \cos\frac{2\pi n_1\xi_1}{l_1}\cos\frac{2\pi n_2\xi_2}{l_2}\cos\frac{2\pi n_3\xi_3}{l_3}, \\ \psi_{n_1,n_2,-n_3} &= a_{n_1,n_2,-n_3} \cos\frac{2\pi n_1\xi_1}{l_1}\cos\frac{2\pi n_2\xi_2}{l_2}\sin\frac{2\pi n_3\xi_3}{l_3}, \end{aligned}\right\} \quad (13\cdot21)$$

etc., and n_1, n_2, n_3 are integers (positive in (13·21)). The factors can be cosines or sines, and (as a convention) we distinguish sine factors by negative values

of n. The condition governing this analysis is that the wave functions defining the elementary states must form a complete set of orthogonal functions capable of representing any distribution of ψ within the block.

Let $$\varpi_\alpha = 2\pi n_\alpha / l_\alpha, \quad \varpi^2 = \varpi_1^2 + \varpi_2^2 + \varpi_3^2. \tag{13·22}$$

Taking ϖ_1, ϖ_2, ϖ_3 as rectangular coordinates, each elementary wave is represented by a point in ϖ-space. The points form a rectangular lattice with intervals $2\pi/l_\alpha$; and their density in ϖ-space is $l_1 l_2 l_3/(2\pi)^3$. By the exclusion principle there are not more than two electrons (and two protons) to an elementary scalar wave.† Hence we shall require $\frac{1}{2}\sigma l_1 l_2 l_3$ waves, which will occupy a minimum volume $\frac{1}{2}(2\pi)^3\sigma$ in ϖ-space. If this volume forms a sphere of radius r, $$\tfrac{4}{3}\pi r^3 = \tfrac{1}{2}(2\pi)^3\sigma. \tag{13·23}$$

The pressure is given by the operator

$$\mathbf{P} = \tfrac{1}{3}(\mathbf{T}_{11} + \mathbf{T}_{22} + \mathbf{T}_{33}) = -\nabla^2/3m \tag{13·24}$$

by (13·12). For an elementary wave ψ_{n_1, n_2, n_3}, this reduces to an eigenvalue

$$P = \varpi^2/3m$$

by (13·21) and (13·22). Hence

$$P l_1 l_2 l_3 = \Sigma\, \varpi^2/3m = \sigma l_1 l_2 l_3 . \overline{\varpi^2}/3m, \tag{13·25}$$

where $\overline{\varpi^2}$ is the mean value of ϖ^2. We have therefore to select the waves which give the minimum value of $\overline{\varpi^2}$. This is obtained by packing them in the sphere of radius r above mentioned. Then

$$\overline{\varpi^2} = \tfrac{3}{5}r^2 = \frac{3}{5}\left(\frac{3}{8\pi}\sigma\right)^{\frac{2}{3}}(2\pi)^2 \tag{13·26}$$

by (13·23). Hence, by (13·25),

$$P_0 = \frac{1}{5}\left(\frac{3}{8\pi}\right)^{\frac{2}{3}}\frac{(2\pi)^2}{m}\sigma^{\frac{5}{3}}. \tag{13·27}$$

In C.G.S. units the factor $(2\pi)^2$ in the numerator is replaced by h^2, our present units being such that $h/2\pi$ is unity. (The coefficient of the momentum operator is $h/2\pi$, not h/π, since ψ is an internal wave function.) The constant m is identified with m_e.

This is called the *ordinary* degeneracy formula. It was first applied by R. H. Fowler‡ in his investigation of the state of matter in white dwarf stars in 1926. But for many years it has been discarded by astronomers in favour of a supposed *relativistic* degeneracy formula. The "relativistic" degeneracy formula appears to be without foundation. The difference is important in the theory of evolution of white dwarf stars; and it was the paradoxical results of the "relativistic" formula, disclosed in an investigation by S. Chandrasekhar,§ which led me to examine its validity.‖

† We take this result from current theory.

‡ *Monthly Notices, R.A.S.* **87**, 114.

§ *Ibid.* **95**, 207.

‖ *Ibid.* **95**, 194; **96**, 20. See also § 13·8.

From a theoretical standpoint it is of greater interest to take the material to be ionised hydrogen, and calculate the whole pressure including proton pressure. The separation of the electron pressure from the proton pressure is somewhat artificial, since the electrons could not exist alone. If σ is the number of scalar particles per unit volume, and m the mass-constant of a scalar particle, i.e. the constant m in (13·12), the minimum pressure is found by the foregoing method to be

$$P_0=\frac{1}{5}\left(\frac{3}{4\pi}\right)^{\frac{2}{3}}\frac{(2\pi)^2}{m}\sigma^{\frac{5}{3}}. \tag{13·28}$$

Suppose that

$$\frac{1}{m}=\frac{136}{\frac{1}{2}m_0}=2\left(\frac{1}{m_p}+\frac{1}{m_e}\right) \tag{13·291}$$

by (12·63). Then (13·28) becomes

$$P_0=\frac{1}{5}\left(\frac{3}{8\pi}\right)^{\frac{2}{3}}(2\pi)^2\left(\frac{1}{m_p}+\frac{1}{m_e}\right)(2\sigma)^{\frac{5}{3}}. \tag{13·292}$$

Since the electron density is now 2σ, this agrees with (13·27) supplemented by a corresponding proton term.

We therefore conclude that the coefficient m for a scalar wave function is as given in (13·291). This is in the system of units with $h=2\pi$. We recall that when the relativistic double valued wave vectors are employed, the units are such that $h=\pi$ (§ 9·6). In that case

$$m=2m_0/136. \tag{13·293}$$

The factor 136 is a transformation factor due to our normal reckoning of mass being determined by double vector wave functions with a phase space of 136 dimensions, whereas the double scalar wave function has a phase space of one dimension. We could eliminate it by assigning appropriate indices to scalar and vector wave functions (8·65). The factor 2 occurs because here the internal coordinates ξ_μ are relative to the centre of mass, whereas in the simple transformation (10·911) ξ_μ is the coordinate of one particle relative to another.

13·3. The Energy of Standing Waves.

The domain of the internal coordinates ξ_μ is three-dimensional, there being no relative time coordinate. Consequently $T_{\alpha\beta}$ for standing waves consists of spatial components only. This is not the whole energy tensor; the remainder is provided by the external wave function which represents the motion of the block as a whole. The rest-mass is contained in the external wave function. In the present coordinate system (chosen so that the block as a whole is at rest) the external energy tensor consists of a single component T_{44}. Since $g^{\alpha\beta}T_{\alpha\beta}=T$, we have in our coordinate system (with real time)

$$T_{44}=T+T_{11}+T_{22}+T_{33}, \tag{13·31}$$

which is also written as

$$\rho=\rho_0+3P. \tag{13·32}$$

Consider now a different set of standing waves in the material of the block, so that the pressure P is changed. One or both of the quantities ρ, ρ_0 must be changed. Let

$$\rho_0 = \rho_0' + \beta P, \tag{13·33}$$

where ρ_0' is constant, and β may be zero or constant or a function of P. The usual point of view is that it is more appropriate to divide the whole energy tensor into a constant external part $T_{44} = \rho_0'$, and group together the part $(P, P, P, (3+\beta)P)$ which comprises the whole effect of the standing waves. To provide for the energy $(3+\beta)P$ we must insert a time factor e^{ikt} in the functions ψ_{n_1, n_2, n_3}, k being dependent on (n_1, n_2, n_3) or equivalently on $(\varpi_1, \varpi_2, \varpi_3)$. Owing to our special choice of axes t is identical with the dynamical coordinate s of the internal state; so it is not so incongruous as it might seem, to mix it with the internal coordinates ξ_μ. There are no boundary conditions for determining k, and it must be determined from the energy density $(3+\beta)P$; but this requires a knowledge of β.

The problem of finding the hamiltonian of the standing waves, i.e. the expression for the energy operator $-i\partial/\partial t$ in terms of the momentum operators $-i\partial/\partial\xi_\mu$, thus resolves itself into determining the constant or function β. We wish to find the change of energy density or mass density δT_{44}, when energy is added to the system in the form of standing waves producing a pressure δT_{11}.

It is commonly taken for granted that the answer to this problem can be checked by observation. No such test is possible; and any answer we may adopt must rest on convention, not observation. Consider a vessel containing gas. I do not doubt that when the pressure is increased by raising the temperature of the gas, the mass (measured in the ordinary way by the acceleration of the vessel under a given applied force) is increased by the amount of the heat energy that is added. Further, if the gas is monatomic, its heat energy is fully represented by standing waves. But for our purpose the experiment is illusory. The integrated pressure of the gas is precisely balanced by the integrated tension of the walls of the vessel. The observed change of mass (integral of δT_{44}) is therefore not associated with any net change of T_{11}, but with a differential effect depending on whether T_{11} is in a gas or a cohesive solid. The complication in the solid is that there are interatomic forces of cohesion, of a type which can only be represented by the use of vector wave functions. We can make an arbitrary addition $\delta T_{44} = \alpha \delta T_{11}$ (α = const.) to the density at each point in the gas and vessel, since it is only possible to test observationally systems in which pressures and tensions balance. Accordingly, $\Sigma \delta T_{11} = 0$; and the addition cancels out on integration over the system.

It may seem more hopeful to examine a steady system without a constraining boundary, e.g. a star cluster, or a star composed of monatomic gas.

But has a star greater or less mass than the sum of the rest masses of its particles? It has more kinetic energy but less total energy. Which of these corresponds to the mass? Or, more precisely, which is represented by the integrated value of T_{44}? The answer to the latter question is *neither*; for T_{44} cannot be integrated in an absolute way in a curved space. In trying to avoid a constraining boundary we have used curvature of space-time (gravitational force and potential energy) to keep the system steady; and the same indeterminacy now appears in the form of non-integrability.

This observational indeterminacy is provided for in the initial formula (13·11). $T_{\mu\nu}$ is indeterminate to the extent of an additive constant $\lambda g_{\mu\nu}$; and the relation of δT_{44} to δT_{11} will depend on the arbitrary choice of a corresponding $\delta\lambda$.

We have therefore to fix the relation between δT_{44} and δP by a convention. For reasons, which we shall presently explain, the convention is taken to be

$$\delta T = 0, \tag{13·34}$$

so that $\beta = 0$. This will fix the change (if any) of the gauge constant λ, and it can therefore be regarded as a gauging equation. It is here asserted only for changes of standing waves; in particular it does not hold for regions in which there is unbalanced angular momentum. But it happens that it includes Maxwellian electromagnetic fields since these have a proper density T which is identically zero.

The invariant T is the *Action* of the material system. The action of electromagnetic waves (but not of aperiodic electromagnetic fields) is zero. Hence (13·34) secures that there is no change of action when radiation is absorbed and converted into standing waves in a material system. It may therefore be interpreted as a *Principle of Stationary Action* for variations of the equilibrium state of matter and radiation. It is in the form of an action principle that the convention (13·34) has become incorporated fundamentally in the current scheme of physics.

We must therefore accept (13·34) as the current convention. Then $\beta = 0$; and, by (13·32) and (13·33), the energy density T_{44} of matter whose internal state is represented by standing waves is

$$\rho = \rho_0 + 3P, \tag{13·35}$$

where ρ_0 is constant.

In the classical theory of gases, and in elementary quantum theory, only the kinetic energy, which is approximately $\frac{3}{2}P$, is considered; and the energy density is taken to be $\rho = \rho_0 + \frac{3}{2}P$. But changes of pressure cannot be produced without changes of the gravitational field; and it is disastrous to introduce relativistic refinements without taking into account the changes of potential energy.

13·4. The Use of an Action Principle.

Wave mechanics is a statistical theory and its results refer primarily to systems in statistical equilibrium. Its procedure is to investigate the possible steady distributions of probability, postulating an ideal environment of the system considered. Its dynamical equations are derived from the condition that a recognisable characteristic of the system remains steady—this being taken as the criterion of a steady state. Herein lies the essence of the statistical method. For the complexions of a system originally regarded as distinct are regrouped according to the values of the selected characteristic; and the probability distribution of its components takes the place of the probability distribution of the original classification. In particular, it is possible to find attributes for which, on the original basis of statistics, some values are infinitely more probable than others—corresponding to singularities in the transformation from the old to the new classification of complexions.

But how can a theory of steady conditions provide anything for observation to get a grip on? Those influences from the external world which reach our senses are due to change and transition.

We have made provision for perturbation of and by these steady systems. The first step is to introduce *steady perturbations*. These are found by treating the perturbed and perturbing system as a combined system in statistical equilibrium, and therefore falling within the scope of the statistical theory. We then derive an equivalent representation as two separate systems each of which is uniformly perturbing the other. The perturbations are expressed as changes of the probability factors attached to the steady states of the two systems.

But it would seem that ultimately there must be some limit to the treatment of phenomena by methods based on the postulate of statistical equilibrium. The universe is far from statistical equilibrium; so that sooner or later we are bound to overstep the limits of the theory. I am not sure that this conclusion is logically sound. According to our usual outlook the universe is far from statistical equilibrium; but it may depend on how we choose the basis of statistical enumeration. The "recognisable characteristic" J of the universe is that it conforms in every detail to our accumulated knowledge of what has actually occurred during a period of a few thousand years. We can at least say that J is constant for changes of the dynamical time coordinate s—except in so far as subjective influences (discovery that certain information is false) may cause "perturbations". But it is perhaps unlikely that the number of symbols $U_1, U_2, \ldots$ commuting with J is sufficient to justify an analogy with the theory of § 9·1.

Be that as it may, physics does not attempt to press the equilibrium theory to such an extreme, but breaks off in a new direction. We must make clear the nature of this fundamental departure.

A simple illustration is afforded by the historical development of the law of gravitation. First the steady states of a combined system—the sun and a planet—were discovered, leading to the formulation of Kepler's laws. Then an equivalent representation as two simple systems, one perturbing the other, was found; and the perturbations were expressed in the form of the inverse-square law of gravitation. The next step was to assume that *the same law of perturbation applies, whether the bodies form a steady system or not.* The Newtonian and Einsteinian pictures of gravitation are such that it seems pedantic to emphasise the arbitrariness of the last step. What possible bearing can the steadiness of the system have on the matter? But our point is (1) that an analogous step must always be taken in developing the general laws of nature from the study of steady states, and (2) it is an exceedingly dangerous kind of generalisation to apply to *statistical formulae.* We know well the many fallacies which have arisen from applying to non-equilibrium distributions the laws found for statistical equilibrium.

We have therefore to recognise that, woven into the method of physics, and forming an indispensable part of it, there is a hypothesis or an assumption or a convention (we leave the appropriate term for further consideration) that results obtained for systems in statistical equilibrium can, in certain circumstances which must be strictly defined, be applied to non-equilibrium systems.

The hypothesis or assumption or convention is the Principle of Stationary Action. This asserts that a certain characteristic (action) of the combined system in statistical equilibrium remains stationary for small deviations from equilibrium. This gives, as it were, a slight play at the joints of our systems, by which we can extricate them from the bondage of statistical equilibrium. After analysing a distribution in statistical equilibrium into a number of separate, but mutually perturbing, systems, we can give those systems a freedom which they did not possess as components of an equilibrium distribution.

The introduction of some such principle is not a wholly arbitrary procedure. Some such "loosening of the joints" is inseparable from the conception of the analysis of a whole into its parts.† It is meaningless to write $a = b + c$ unless we contemplate the possibility that b may have a significance when c is not added to it. As a condition for detaching b from c, we must recognise a definite distribution of the characters of a between b and c. Thus an important aspect of the principle of stationary action is a *localisation* of the characteristics—energy, spin, etc.—of the combined system.

Let us return to the problem of determining observationally the change

† Cf. the introduction of an infinitesimal element of relative time dt_r in § 12·3. It is there postulated that the probability distribution of the combined system in phase space is stationary for such a variation.

of energy with pressure of a gas. To render the conditions static and thereby amenable to treatment by the statistical theory which contains our fundamental definitions, it was necessary to enclose the gas in a vessel. But the enclosure introduced compensating tensions which frustrated us. Having introduced the envelope, we were unable to detach the gas from its envelope. As an alternative we may consider a small volume of gas in the interior of a star, and seek to determine by ideal observations how its energy would be changed by a change of pressure. The question is absurd; we cannot change the pressure at one point in a star (in steady conditions) without altering the whole star. Observation will not tell us what part of the whole change of energy is located in the particular volume considered.

To attach an observational meaning to a *local* association of energy density and pressure, we must be able to produce a change whose effects are confined to the locality. Such a change is provided by the conversion of radiation into molecular motion of matter. According to (13·34), if the radiant energy is converted wholly into standing waves, as in an ideal monatomic gas, there is no change of energy density or pressure;† so that the effects are confined to the locality. In particular the gravitational field emanating from the region is unaltered, so that there is no cause of readjustment of the matter outside. But there is the preliminary objection that the existence of radiation in any other than its equilibrium proportion presupposes a highly disturbed state of the star. Thus even the conversion of radiation into molecular motion is not a strictly local phenomenon. Its effects are local, but its causes are not local. It cannot occur independently of a general settling down of the star, implying similar conversions in other regions.

It is here that the principle of stationary action, which must now have the definite form (13·34), steps in. It asserts that $\delta T = 0$ applies to the local conversion of radiation into material energy, independently of the conversions occurring in other parts of the star. Although the rest of the star could not actually be in statistical equilibrium, it may be treated as if it were in statistical equilibrium; because the property with which we are concerned is stationary for small deviations from statistical equilibrium.

The mathematical form of the action principle shows quite explicitly that it is a means of localising the characteristics of the universe. The quantity to be varied is an integrated quantity covering a large volume; the quan-

† This appears to be inconsistent with the elementary formulae, which predict a change of pressure if the volume is unchanged; but in our problem we have to admit whatever change of volume is necessary in order that the effect of conversion may be strictly localised, i.e. that there may be no change of the gravitational field outside the region. But a real change of volume would displace the surrounding matter outwards or inwards; the change must therefore be represented as a change of *reckoning* of volume, implying a change $\delta\lambda$ of the gauge-constant.

tities determined or defined by the variation are functions of position—so that each minute part of the system is described separately. Localisation is an artificial conception in an interrelated universe, where the influence of any one part extends through the whole. To detach one part from the rest is impossible. The action principle reassures us by asserting that (with suitable safeguards) a mild impossibility is a permissible idealisation. Its ill effects are of the second order of small quantities.

Like many of the "principles" in science, the genesis of the action principle is that, having realised that we ought not to make a tempting assumption, we erect a principle to say that we may make it. I shall not attempt to defend the principle of stationary action as a physical assumption; there is no need to do that. Its defence is that it is the basis of our definitions. A principle of localisation must precede the definition of any localised entity, e.g. the $g_{\mu\nu}$ or $F_{\mu\nu}$ of field theory. By definitions I here mean, not the mathematical definitions as symbols in a deductive theory, but the definitions by which they are recognised and measured in observational science. The action principle is not a physical hypothesis; it is a means of defining localised quantities. Thus in the application in § 13·3 we had an undefined disposible constant λ. By adopting the action principle (13·34) we remove the arbitrariness of $\delta\lambda$, and changes of energy and pressure are defined precisely.

In considering the action principle as a vehicle of definition, it is desirable to distinguish between:

(*a*) *Weak action principles.* Referring to conditions in empty space, with or without weak electromagnetic fields.

(*b*) *Strong action principles.* Referring to conditions in continuous matter, intense electromagnetic fields, the interior of a nucleus or electron, etc.

The difference arises because, although an investigator has a certain amount of freedom in adopting definitions of the terms which he employs, he must have regard to established usage. In regard to (*a*) established usage dictates the form of the action principle. There is no difference of opinion as to how the kernel of the action invariant is constituted, to the order of accuracy required in weak conditions; for it has to lead to definitions already recognised. In regard to (*b*) an investigator has entire freedom in his choice of action invariant and consequent definitions of $g_{\mu\nu}$ and $F_{\mu\nu}$, provided only that it converges to the weak action invariant as weak conditions are approached. Current literature shows that full advantage is taken of this freedom!

Observational measurements are only made under weak conditions. The "distance between two points" in field-free conditions is what it is determined to be by measurement; the distance between two points in an intense magnetic field is whatever (within reason) a theorist chooses to call it. No

experimental physicist would attempt accurate measurement of distance in an intense magnetic field; he would mistrust his scales or other apparatus. In practice he does not measure the curvature of the track of an electron in an intense magnetic field; he measures the curvature of a track on a photographic plate placed outside the field. Intervals, and the corresponding $g_{\mu\nu}$, are only defined observationally in field-free conditions, or when theory (that is to say, the accepted action invariant) assures us that the electromagnetic field is not strong enough to introduce sensible error. When once we begin to suspect that our measured distances need correction, we turn to theory to tell us what the measures ought to have been. Naturally the corrected measures will confirm the theory used for determining the corrections, whatever form that theory may take.

The literature of mathematical physics abounds in proposals of universal action principles covering strong as well as weak conditions. I think it is not unfair to summarise the majority of such proposals as follows:

(1) A geometry based on a new set of axioms is outlined.

(2) A fundamental invariant of the new geometry is selected as action invariant.

(3) It is shown that the action principle leads to the usual equations for weak fields.

(4) Second order terms, sensible in strong electromagnetic fields, are found, which it is suggested may provide an observational test of the theory.

(5) A hope that something further may come of it.

Our comment is: (1) We may choose any kind of space we please for the purpose of graphical representation of physical quantities without committing ourselves to anything.† (2) The quantities $g_{\mu\nu}$, $F_{\mu\nu}$ are to be used in strong conditions in which there is no agreed definition as to how they are to be measured. The action invariant embodies the definition selected in the new theory. (3) This shows that the definition is not in conflict with any established usage. (4) The result of the test will inevitably be a triumphant verification of the new theory, provided that the observations are correctly reduced. By *correctly* we mean that, for measurements with apparatus actually located in the intense fields, the readings of the apparatus (which are, of course, affected by the field) are corrected to accord with the readings of ideal apparatus which conforms to the equations of the new theory; and for measurements made outside the fields, the inference as to what is happening within the field is calculated according to the equations of the new theory.

The trouble about unified field theories is that there are so many of them, and all of them are right. The various action principles are various plans of

† *Mathematical Theory of Relativity*, § 83.

localising the characteristics of the interrelated conditions which the universe presents to us. The procedure of localising an entity is inseparable from the procedure of defining a local entity. It all reduces to a question of definitions. The difference between the rival theories really rests on (5)—as to which opinions will naturally differ.

13·5. Potential Energy.

Let us return to the problem of the rectangular block. In order to avoid inessential complications, we shall at first suppose the density to be very low, so that collisions of the particles can be neglected. From one aspect the block is occupied by the set of standing waves introduced in § 13·2. From another aspect it is occupied by n elementary particles moving in diverse directions; these are represented by progressive waves which combine into an n-tuple wave function. I think it will save much confusion of thought if we notice that the standing wave function represents a system in equilibrium, and the n-tuple wave function represents a rapidly dispersing system. If we insert in the n-tuple wave function a later value of the time, it gives the distribution which would be reached if each particle continued to move with its present momentum.

Let us then take a somewhat modified n-tuple wave function Ψ, which at every time t represents the n particles occupying the block at that time. This will represent a statistically steady system, just as the standing wave function does. Every time a particle crosses the boundary Ψ changes discontinuously.† We must replace the discontinuities by continuous change; otherwise the function is not differentiable and is useless for computing energy. But the energy $-i\partial/\partial t$ of this smoothed wave function will be altogether different from that of the instantaneous n-tuple wave function.

The general effect of smoothing out the discontinuities can be seen by considering the combined wave function of the rectangular volume A and the surrounding matter B which holds it in equilibrium. The steady state (represented by standing waves) results from the fact that on the average for every particle passing from A to B, a similar particle passes from B to A. In the combined wave function of A and B this is merely a nominal interchange. The coordinates x_p of the pth particle of system A have become inappropriate to system A, and the coordinates x_q' of the qth particle of system B have become inappropriate to system B; we therefore re-label the particle at x_p as the qth particle of system B, and the particle at x_q' as the pth particle of system A. By the well-known Fermi-Dirac rule this interchange of labels reverses the sign of the wave function—in this case the

† Since we are supposed to have approximate knowledge of the momenta of the particles, we cannot know their positions exactly. Thus the time of crossing the boundary is indefinite to a slight extent; and the change of Ψ though erratic is not strictly discontinuous.

combined wave function of A and B. In time-averaging the succession of reversals of sign is replaced by a continuous factor $e^{ip_v t}$, which reverses the sign at regular intervals π/p_v. When we form the energy $-i\partial/\partial t$, the factor $e^{ip_v t}$ gives an additional energy p_v. This interchange energy does not appear in the instantaneous n-tuple wave function which represents a dispersing system; but it belongs inalienably to the steady system represented by the modified function Ψ or alternatively by standing wave functions.

Interchange energy is evidently potential energy. (I may add that so far as we know all potential energy is interchange energy.) The system occupying the rectangular block may be kept in equilibrium, either by interchange of particles at the boundary, as described above, or by a force exerted by the matter beyond the boundary turning back the particles as they reach it. But the two processes are the same. For, since the particles are indistinguishable individually, it is meaningless to discriminate between an old particle being turned back and a new particle coming in. Thus the interchange effect and the force are identical. If we happen to take one view we call the energy interchange energy; if we take the other view we call it potential energy in the field of force.

The rectangular block, which we have been contemplating, must not be too small to be treated macroscopically. In many applications the free path of the particles will be small compared with l_1, l_2, l_3. The collisions are then the main instrument in preventing the particles from straying outside the block; the reduced number which hover about the boundary are deemed to be turned back by interchange as before. We shall find in Chapter XV that the electrical forces controlling the encounters of protons or electrons are attributable to interchange. For our purposes it is unimportant whether the potential energy included in T_{44} is due to collision interchange or to boundary interchange. The formal difference is that collision interchange is provided for in the equations and definitions of electromagnetic theory, so that it is an intrinsic property of electrons and protons as currently defined; boundary interchange is an averaging adjustment which enables us to treat a dispersing system as a static system. We may say that when the free path is small compared with a macroscopic volume element, the volume element *is genuinely* in equilibrium; when the free path is long, it *can be treated as* in equilibrium.

When the velocities of the particles are not too great, and at the same time the density is too low to give rise to degeneracy, the classical theory applies and the density of the kinetic energy is $\frac{3}{2}P$. To make up the total $\rho_0 + 3P$, the potential energy must be $\frac{3}{2}P$. Thus the kinetic and potential energies are equal. The fact is that the waves represented by scalar wave functions are the ordinary elastic vibrations or sound waves of the material. The only

point at which we go beyond classical dynamics is when we apply the exclusion principle.

The exclusion principle sets an upper limit to the amplitude of any particular wave, such that its energy cannot exceed one quantum. In non-equilibrium conditions the amplitude of a particular wave can greatly exceed this limit, e.g. when an organ pipe is sounding its fundamental note. This emphasises the fact that there is no justification for applying the exclusion principle in non-equilibrium conditions. It is intimately connected with the interchange forces which maintain equilibrium, as will be seen in Chapter XV.

In extreme conditions the kinetic and potential energies are no longer equal. For constant density, it appears that the energy $3P$ is wholly potential at absolute zero (complete degeneracy) and tends to become wholly kinetic at very high temperatures. We should first make sure that this separation into kinetic and potential energy has an observational meaning. Certain phenomena, e.g. the rate of disruption of nuclei by protons, evidently depend on the kinetic energy of the elementary particles (not the composite average particles); and since the rate of disruption increases rapidly with the speed of the protons, the distribution law of kinetic energies of the individual protons is involved, and doubtless will in due time be ascertained from observations of this kind. The problem of finding this distribution law theoretically should be soluble; but apparently it has not yet been solved. The individual protons and electrons correspond to progressive waves; the present investigation treats only the composite particles represented by standing waves, and is not relevant.

Reference must be made to the current formulae (associated with the "relativistic" degeneracy theory) which profess to give the distribution law of the kinetic energies and momenta of the individual protons and electrons at all temperatures. So far as I know, they may be right at high temperatures; but it is impossible to trust them, not only because they rest on a false conception of relativity, but because they are clearly wrong at low temperatures. In these investigations the minimum pressure P_0 which corresponds to absolute zero is attributed to kinetic energy of the elementary particles. If that were so the protons would, if the density were great enough, still have large velocities at absolute zero, and continue to disrupt the atomic nuclei. This surely is a contradiction of thermodynamical principles. Our own investigation does not discriminate between kinetic and potential energies; but coupled with the general physical principle that processes of the nature of ionisation or transmutation must cease at absolute zero, we infer that the energy $3P_0$ must be wholly potential.

It may be urged in defence of the current formulae that the energies and momenta to which they refer are not to be taken as the energies and momenta for the purpose of calculating collision effects—that they are not

to be interpreted according to the classical conception of motion. To this we reply: (1) Then the distinction between kinetic and potential energy loses all observational meaning; and we lose nothing by treating them together in our own theory. But the question still remains unanswered, What is the distribution law of the energies and momenta concerned in collision effects? (2) Our criticism of the current theory is that it *does* interpret the energy $\frac{3}{2}P_0$ according to the classical conception of motion; it treats the corresponding velocity or momentum as a vector to which Lorentz transformations may be applied, and thereby introduces change of mass with velocity. If the particle has one velocity for the purpose of Lorentz transformations and another velocity for the purpose of collision phenomena, what becomes of the principle of relativity?

13·6. Transition to Vector Wave Functions.

For the problem of the rectangular block we have used a form of wave mechanics much simpler than that which has been developed in previous chapters for the treatment of microscopic problems. If it is asked why the simple form does not apply to an atom, we answer that it does not even apply to a macroscopic distribution of matter in a smooth *spherical* vessel. In addition to the modification of the elementary orthogonal functions ψ_{n_1,n_2,n_3} to correspond to a spherical boundary, a new feature appears. There are steady states in which the distribution has a resultant angular momentum.

With a spherical boundary unidirectional motion is no longer inconsistent with a steady state. In rectangular coordinates there must be on the average as much motion in one direction as in the other direction of ξ_α, and the state of motion is accordingly represented by standing waves; but in polar coordinates it is not necessary that there should be as much motion in one direction as in the other direction of θ, so that we may have progressive waves in an angular coordinate θ. Such waves are not essentially different from the progressive waves representing a particle moving freely in space, i.e. revolving unidirectionally about the centre of curvature of space.

The advent of progressive waves into a problem depends on the existence of relativity transformations of the coordinates, which occasion a degeneracy of the ordinary steady state solutions. In the problem of the rectangular block there are no relativity transformations of the internal coordinates ξ_μ; taking account of boundary conditions, no other axes are "equivalent" to those which are parallel to the edges of the block. In the spherical problem all orientations of the internal rectangular axes are equivalent; so that rotations in the three coordinate planes are relativity rotations. The time direction must be chosen so that the spherical boundary is at rest; it is

therefore still unique, and there are no Lorentz transformations of the internal coordinates.

The machinery for dealing with relativity transformations has been developed in our early chapters; and in order to provide for the relativity transformations of the internal coordinates we must introduce it into the internal wave functions. But for three-dimensional relativity it is unnecessary to resort to Dirac's fourfold wave vectors. We can take ψ, ϕ to be wave vectors with two components (2-vectors), and associate the coordinate planes with Pauli matrices ζ_1, ζ_2, ζ_3. As explained in § 3·8, ζ_1, ζ_2, ζ_3, i form a minor complete set, and are a simplified representation of the symbols E_{12}, E_{23}, E_{31}, E_{16} when the other symbols are not being used.

Considering an elementary scalar wave, let σ be the particle density, or the probability that the particle belonging to the wave is within a unit volume at the point considered. If ψ, ϕ are normalised in accordance with the exclusion principle (13·144), $\sigma=\psi\phi$. Thus scalar wave functions relate to the special case in which the stream strain vector $S=\psi\phi^*$ consists of a single algebraic component σ. But when ψ and ϕ are 2-vectors, an algebraic quantity σ cannot be a simple product $\psi\phi^*$; σ is the sum of two pure components

$$\tfrac{1}{2}\sigma(1+i\zeta_\mu), \quad \tfrac{1}{2}\sigma(1-i\zeta_\mu), \tag{13·61}$$

where ζ_μ is the Pauli matrix associated with an arbitrary plane. The two expressions in (13·61) are idempotent when normalised (so that the halfspur $=\frac{1}{2}$); hence they can be factorised (§ 5·6).

Accordingly the "particle" corresponding to an elementary scalar wave is replaced by two sub-particles with the non-algebraic stream vectors (13·61). The components $\frac{1}{2}\sigma$ and $\frac{1}{2}i\sigma\zeta_\mu$ correspond to energy and angular momentum. Thus the sub-particles have opposite spin in the ζ_μ plane. When the boundary conditions are such that the angular momentum in a steady state is zero, the oppositely spinning particles are constrained to occur together; they can be regarded as together constituting an indivisible unit—a scalar particle: and we are only concerned with the sum of their strain vectors which, being an algebraic quantity, is represented by scalar wave functions. But when, as in the spherical problem, integrals of angular momentum exist, the oppositely spinning particles have independent probability distributions and are represented by separate 2-vector wave functions.

The analysis of a system into elementary states has to be considered in conjunction with the perturbations anticipated. The elementary state is a unit which requires a separate probability factor in perturbation theory. In earlier chapters we have had in mind well-isolated systems, whose reaction with other systems is limited to weak or occasional perturbations. But in these macroscopic problems the "system" described by the wave functions

is usually a volume more or less arbitrarily carved out from its surroundings, and the viscous forces† at the boundary constitute a strong permanent perturbing influence. We have different elementary states of a gas contained in a spherical vessel according as the sphere is rough or smooth, because in the rough sphere the permanent perturbation is such as to prevent any persistence of probability in one direction of spin as compared with the other; there is no occasion to give them separate probability factors.

The choice of elementary states depends on the *approximate* rather than on the *exact* boundary conditions. A vessel cannot be perfectly smooth; and if we paid attention to the exact boundary conditions, we should conclude (correctly) that the only state of statistical equilibrium was one of zero angular momentum. But the gas may take a long while to reach this state; and we learn more by utilising the complete series of states in an ideally smooth vessel, treating the small friction as a perturbation which slowly transfers the excess probability of one direction of spin to the opposite direction.

The problem of a smooth spherical vessel is a half-way stage between the rectangular block and a particle moving freely in space-time. For the free particle the relativity transformations are extended to include Lorentz transformations. This additional degeneracy must be provided for by a second set of Pauli matrices θ_μ, commuting with the ζ_μ. The two components of opposite spin $\frac{1}{2}\sigma(1 \pm i\zeta_\mu)$ are in turn analysed into two components of opposite electrical sign

$$\tfrac{1}{4}\sigma(1 \pm i\zeta_\mu)(1 \pm i\theta_\mu). \qquad (13{\cdot}62)$$

The wave vectors are now 4-vectors, and the double set of Pauli matrices is more conveniently replaced by a set of Dirac matrices with the same commutative relations (§ 3·8). The four components (13·62) then become the spectral components given in (6·64).

It is of interest to consider whether (13·62) has any application to macroscopic systems. We might start with θ_μ instead of ζ_μ, and divide the scalar density σ into two components of opposite electrical sign $\frac{1}{2}\sigma(1 \pm i\theta_\mu)$. The scalar wave function postulates that these are constrained to have equal probability so that they need not be considered separately. Can the boundary conditions be modified so that they have independent probabilities? I think so. The necessary condition is that the system shall be *insulated*. With an insulated boundary, there are steady states having an excess of components of one electrical sign; just as with a smooth spherical boundary there are steady states having an excess of components with one direction of spin.

† The boundary pressure (interchange force) is allowed for by including potential energy in the system, so that we avoid treating it as an extraneous perturbation.

If the substance is permanently magnetisable it is also possible to have steady states with a preponderance in one direction of the magnetic stream vector.

These states of electric or magnetic excess may also be induced by permanent perturbation from outside, i.e. by steady electric or magnetic fields. Thus the general problem of statistical equilibrium of a macroscopic system, when the angular momentum is not constrained to be zero and electric and magnetic fields are not excluded, will involve Dirac's vector wave functions.

When the spins and charges are not balanced, the state of the material is only imperfectly described by the ordinary energy tensor. For a full description we require the Riemann-Christoffel matrix (§ 11·4). But it should be remembered that the electric and magnetic fields which occur in practice in macroscopic systems correspond to an entirely trivial excess of components of one sign. So that whilst these phenomena are important on their own account, there is very little to represent them in the statistics. It is therefore more suitable for ordinary purposes to represent the electromagnetic field separately by potential theory, rather than to merge it in the general statistical formulation where it would be almost lost.

If we apply the energy operator (13·12) to vector wave functions, the covariant differentiation will introduce matrices (§ 8·3). Each component of the energy tensor therefore becomes a matrix, and we obtain altogether 256 components as when the Riemann-Christoffel matrix is used. It appears therefore that the theory of the energy of distributions which involve vector wave functions, treated in Chapter XI, could be reached by starting with the energy operator (13·12).

13·7. Origin of the Energy Operator.†

In this section we shall obtain the formal connection between the energy tensor

$$T_{\alpha\beta} = -(1/8\pi\kappa)\{G_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}(G - 2\lambda)\}$$

and the energy operator

$$\mathbf{T}_{\alpha\beta} = -m^{-1}\delta^2/\delta x_\beta \delta x_\alpha .$$

By contracting the energy tensor we have

$$T = (1/8\pi\kappa)(G - 4\lambda).$$

Using this to eliminate λ, we obtain‡

$$T_{\alpha\beta} - \tfrac{1}{4}g_{\alpha\beta}T = -(1/8\pi\kappa)(G_{\alpha\beta} - \tfrac{1}{4}g_{\alpha\beta}G). \qquad (13{\cdot}71)$$

† As this section is occupied with ordinary tensor calculus the summation convention is used unrestrictedly.

‡ I was led to consider $G_{\alpha\beta} - \frac{1}{4}g_{\alpha\beta}G$ through correspondence with Dr J. Ghosh.

Consider the gauge-invariant curvature tensor†

$$*G_{\alpha\beta}=G_{\alpha\beta}-3(\kappa_\alpha)_\beta+(\kappa_\beta)_\alpha-g_{\alpha\beta}(\kappa^\epsilon)_\epsilon-2\kappa_\alpha\kappa_\beta+2g_{\alpha\beta}\kappa_\epsilon\kappa^\epsilon, \quad (13{\cdot}721)$$

which is invariant for the transformation

$$g_{\alpha\beta}'=\beta^2 g_{\alpha\beta}, \quad \kappa_\alpha'=\kappa_\alpha+\partial\theta/\partial x_\alpha, \quad (13{\cdot}722)$$

where β is any scalar function of the coordinates, and

$$\theta=\log\beta. \quad (13{\cdot}723)$$

Multiplying (13·721) by $g^{\alpha\beta}$, we obtain

$$*G=G-6(\kappa^\epsilon)_\epsilon+6\kappa_\epsilon\kappa^\epsilon. \quad (13{\cdot}724)$$

Hence
$$G_{\alpha\beta}-\tfrac{1}{4}g_{\alpha\beta}G=*G_{\alpha\beta}-\tfrac{1}{4}g_{\alpha\beta}*G+2(\kappa_\alpha\kappa_\beta-\tfrac{1}{4}g_{\alpha\beta}\kappa_\epsilon\kappa^\epsilon)$$
$$+2\{(\kappa_\alpha)_\beta-\tfrac{1}{4}g_{\alpha\beta}(\kappa^\epsilon)_\epsilon\}+F_{\alpha\beta}, \quad (13{\cdot}725)$$

where $F_{\alpha\beta}=(\kappa_\alpha)_\beta-(\kappa_\beta)_\alpha$.

Take $\kappa_\alpha=0$ initially; and let Δ denote an increment produced by the gauge transformation. Then, since $*G_{\alpha\beta}-\frac{1}{4}g_{\alpha\beta}*G+F_{\alpha\beta}$ is invariant for the transformation, (13·725) gives

$$\Delta(G_{\alpha\beta}-\tfrac{1}{4}g_{\alpha\beta}G)=2(\theta_\alpha\theta_\beta-\tfrac{1}{4}g_{\alpha\beta}\theta_\epsilon\theta^\epsilon)+2\{(\theta_\alpha)_\beta-\tfrac{1}{4}g_{\alpha\beta}(\theta^\epsilon)_\epsilon\}, \quad (13{\cdot}731)$$

where θ_α denotes $\partial\theta/\partial x_\alpha$.

If δ denotes covariant differentiation

$$\frac{\delta^2}{\delta x_\beta\delta x_\alpha}(\log\beta)=\frac{1}{\beta}\frac{\delta^2}{\delta x_\beta\delta x_\alpha}\beta-\frac{1}{\beta^2}\frac{\delta\beta}{\delta x_\beta}\frac{\delta\beta}{\delta x_\alpha}$$

so that
$$(\theta_\alpha)_\beta=\beta^{-1}(\delta^2/\delta x_\beta\delta x_\alpha)\beta-\theta_\beta\theta_\alpha. \quad (13{\cdot}732)$$

Let
$$-\frac{1}{4\pi\kappa}\frac{\delta^2}{\delta x_\beta\delta x_\alpha}=\mathbf{T}_{\alpha\beta} \quad (13{\cdot}74)$$

so that
$$(\theta_\alpha)_\beta+\theta_\alpha\theta_\beta=-4\pi\kappa.\beta^{-1}\mathbf{T}_{\alpha\beta}\beta \quad (13{\cdot}751)$$

and, by contraction,
$$(\theta^\epsilon)_\epsilon+\theta^\epsilon\theta_\epsilon=-4\pi\kappa.\beta^{-1}\mathbf{T}\beta. \quad (13{\cdot}752)$$

Then, by (13·71), (13·731), (13·751), (13·752),

$$\Delta(T_{\alpha\beta}-\tfrac{1}{4}g_{\alpha\beta}T)=\beta^{-1}(\mathbf{T}_{\alpha\beta}-\tfrac{1}{4}g_{\alpha\beta}\mathbf{T})\beta. \quad (13{\cdot}753)$$

We now introduce wave functions

$$\psi=\beta, \quad \chi=\beta^{-1}, \quad (13{\cdot}76)$$

representing an addition to, or modification of, the distribution whose energy tensor is $T_{\alpha\beta}$. Our result shows that instead of expressing the modification as a change of curvature of space-time (which would involve altering the metrical tensor $g_{\alpha\beta}$ to $\beta^2 g_{\alpha\beta}$), we can express the change of $T_{\alpha\beta}-\frac{1}{4}g_{\alpha\beta}T$ as the expectation value of an operator $\mathbf{T}_{\alpha\beta}-\frac{1}{4}g_{\alpha\beta}\mathbf{T}$, the operator being defined by (13·74).

This is the most primitive connection between the wave functions and

† *Mathematical Theory of Relativity*, equation (87·5); the suffixes after the brackets denote covariant differentiation. Although this tensor is generally studied in connection with Weyl's geometry, we do not here apply it in that way. We use only the well-known analytical property of invariance of the tensor.

operational momenta of wave mechanics and the curvature tensors which represent momenta, etc., in general relativity theory.

Equation (13·753) can be resolved into separate equations:

$$\Delta T_{\alpha\beta} = \chi \mathbf{T}_{\alpha\beta}\psi, \quad \Delta(g_{\alpha\beta}T) = g_{\alpha\beta}.\chi\mathbf{T}\psi, \tag{13·77}$$

provided that the change $\Delta\lambda$ of the arbitrary constant λ is suitably chosen. The two equations are consistent since

$$\Delta(g_{\alpha\beta}T) = \Delta(g_{\alpha\beta}g^{\epsilon\zeta}T_{\epsilon\zeta}) = g_{\alpha\beta}g^{\epsilon\zeta}\Delta T_{\epsilon\zeta} = g_{\alpha\beta}g^{\epsilon\zeta}.\chi\mathbf{T}_{\epsilon\zeta}\psi,$$

$g_{\alpha\beta}g^{\epsilon\zeta}$ being unaltered by change of gauge. The second equation of (13·77) is therefore derivable from the first.

The wave functions employed in practical applications of wave-mechanics are an adaptation of these primitive wave functions to specially simple conditions. In particular a modified energy operator is used, defined by

$$\mathbf{T}_{\alpha\beta} = -\frac{V}{4\pi\kappa}\frac{\delta^2}{\delta x_\beta \delta x_\alpha}, \tag{13·78}$$

whose eigenvalue is the *integrated* energy over a three-dimensional domain of volume V. Instead of ψ, χ we take two equal wave functions

$$\psi = \phi = \beta.$$

The product $\psi\phi$ bears the same kind of relation to $\psi\chi$ that a strain vector bears to a space vector. Instead of (13·77) we have

$$V.\Delta T_{\alpha\beta}.\phi\psi dV = \phi\,\mathbf{T}_{\alpha\beta}\psi dV.$$

If now ϕ and ψ are normalised so that $\int\phi\psi dV = 1$, we have

$$V.\Delta T_{\alpha\beta} = \int \phi\,\mathbf{T}_{\alpha\beta}\psi.dV$$

and $V.\Delta T_{\alpha\beta}$ is regarded as the energy of a single particle distributed throughout V with a probability density $\phi\psi$.

Further light on the connection between wave functions and gauge transformations will be found in § 14·1, where a special case is treated.

Although we do not normally employ Weyl's representation of the electromagnetic field in wave mechanics, it is of some interest to extend the foregoing investigation to his theory. The difference is that we can no longer put $\kappa_\alpha = 0$ initially, since the electromagnetic force $F_{\alpha\beta}$, which is invariant for the gauge transformation, is represented by curl κ_α. The term

$$2(\kappa_\alpha\kappa_\beta - \tfrac{1}{4}g_{\alpha\beta}\kappa_\epsilon\kappa^\epsilon)$$

in (13·725), being non-linear, will give cross terms

$$2(\kappa_\alpha\theta_\beta + \kappa_\beta\theta_\alpha - \tfrac{1}{2}g_{\alpha\beta}\kappa^\epsilon.\theta_\epsilon)$$

or, in operational form,

$$2(\kappa_\alpha\partial/\partial x_\beta + \kappa_\beta\partial/\partial x_\alpha - \tfrac{1}{2}g_{\alpha\beta}\kappa^\epsilon\partial/\partial x_\epsilon), \tag{13·79}$$

representing a mutual energy of the waves and the electromagnetic field, which is not included in (13·74).

13·8. The Degeneracy Formula.

Since the formula (13·27) for the minimum pressure differs from the so-called relativistic degeneracy formula which has been widely but uncritically accepted, we shall here explain the origin of the difference.

If a particle of proper mass m has a velocity vector dx_μ/ds, its "integrated energy tensor" is defined to be

$$U^{\mu\nu} = m\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}. \tag{13·811}$$

Suppose that there is a discrete set of values of $U^{\mu\nu}$, and that there are n_r particles with energy $U_r^{\mu\nu}$ in a three-dimensional volume V. A fractional value of n_r is to be interpreted as a probability. The resultant energy tensor (reckoned as usual per unit volume) is

$$T^{\mu\nu} = \Sigma_r n_r U_r^{\mu\nu}/V. \tag{13·812}$$

Let
$$N_r = n_r\,(dt/ds)_r. \tag{13·82}$$

Then N_r/V is the proper density of distribution of the particles with suffix r as contrasted with the relative density n_r/V. The energy tensor can also be written

$$T^{\mu\nu} = \frac{1}{V}\Sigma N m\frac{dx_\mu}{ds}\frac{dx_\nu}{dt} = \frac{1}{V}\Sigma N M\frac{dx_\mu}{dt}\frac{dx_\nu}{dt}, \tag{13·83}$$

where $M = m\,dt/ds$.

The exclusion principle provides an upper limit to the number n_r or N_r. The difference between the two theories is

$$\left.\begin{aligned}&\text{Present theory } n_r \leqslant 1,\\ &\text{Current theory } N_r \leqslant 1.\end{aligned}\right\} \tag{13·84}$$

It will be seen that, if $V = 1$, the maximum contribution of any one state to $T^{\mu\nu}$ is on the present theory $m\,(dx_\mu/ds)\,(dx_\nu/ds)$, and on the current theory $m\,(dx_\mu/ds)\,(dx_\nu/dt)$; or in terms of momenta the contributions are $p^\mu p^\nu/m$ and $p^\mu p^\nu/p_0$. In particular the contributions to the energy are, respectively, ${p_0}^2/m$ and p_0.

It is clear that N_r has been used in current theory under the impression that (13·812) is a tensor equation. If each term on the right-hand side were a tensor, n_r/V would be invariant for Lorentz transformations, and N_r would be invariant. But the eigenvalues $U_r^{\mu\nu}$ of a tensor operator do not constitute a tensor, and (13·812) is not transformable.

When once it is realised that the calculation of P_0 is concerned with the random internal motions relative to the centre of gravity of the whole block, so that the time-axis is prescribed, it is difficult to see any reason at all for the introduction of N_r. For standing waves the momentum does not reduce to an eigenvalue, so that the number N_r does not exist.

Since $N_r m = n_r M$, the substitution of N_r for n_r is from one point of view equivalent to introducing change of mass with velocity. The pressure in a

star is of order of magnitude $\rho\phi$, where ϕ is the gravitation potential measured from the boundary. If all the particles are of the same mass, v^2 is of the same order as ϕ; and the change of mass with velocity $\frac{1}{2}mv^2/c^2$ is of the same order of magnitude as the gravitational potential energy $m\phi/c^2$, whose effect on the mass is neglected. It is, however, assumed in the current theory that, although v^2 is of order ϕ for protons and ions, it is very much higher for electrons, which thereby contribute the bulk of the pressure. But there is no reason to suppose that in the degenerate state the electrons have higher speeds than the protons. Equipartition of energy refers only to transferable energy.

Investigators seem to have been misled by trusting to the classical picture of moving particles. When a particle is forced up into a state of high energy by the occupation of the lower states, the current theory pictures the energy in the classical way as translation with high velocity. It accepts this so literally that it supposes that the particle could be reduced to rest by a Lorentz transformation, and thereby calculates its change of mass with velocity. But non-transferable energy cannot logically be represented that way. We take it that the degeneracy energy is potential energy. There is then no correction for change of mass with the (non-existent) velocity; and the exclusion principle must be $n_r \leqslant 1$, since N_r is undefined.

The exclusion principle was first recognised in the atom where it applies to the internal motion of a system in a steady state. In generalising it to macroscopic systems it would be unsafe to disregard this restriction. In any case the problem to which we here apply it relates to an internal system in a steady state. If we are not obsessed with the idea that the formulae must somehow be capable of extension to dispersing systems represented by plane progressive waves, or that they must be invariant for transformations of axes which would change the nature of the problem, the procedure is straightforward and unambiguous. The steady state distribution is represented by standing waves. The choice between n_r and N_r is settled by the fact that N_r is not only irrelevant but non-algebraic. As shown in § 13·5 the steady condition involves a potential energy due to interchange, and for complete degeneracy the energy is wholly potential. Finally, we shall be able to test the theory in Chapters XIV and XVI, by using it to calculate the constant of gravitation and the cosmical constant.

Originally my choice of n_r was derived from the investigation in § 10·9, where it was shown that the relative energy of a particle is not p_0 but $\alpha(p_1^2+p_2^2+p_3^2)/m$. Since the law connecting pressure and density in a gas depends only on the internal motions, which must be separated from the mass motion of the gas as a whole, I could not feel satisfied with the current theory which took the energy contribution due to one occupied state to be p_0.

The present theory gives $P_0 \propto \sigma^{\frac{5}{3}}$ at all densities. The current theory gives $P_0 \propto f(\sigma)$, where $f(\sigma)$ changes gradually from $\sigma^{\frac{5}{3}}$ to $\sigma^{\frac{4}{3}}$ as the density increases. This is especially important in the theory of white dwarf stars (in which the degeneracy pressure becomes large), because it can be shown rigorously that the pressure in a star cannot exceed $B\rho^{\frac{4}{3}}$, where B is a constant depending only on the star's mass.† The curve $P_0 = K\sigma^{\frac{5}{3}}$ giving a minimum pressure must, as σ and ρ increase, eventually cross the curve $P = B\rho^{\frac{4}{3}}$ giving a maximum pressure. The crossing point is an extreme upper limit to the density in a star of given mass; this limit is of order 10^7 gm. cm.$^{-3}$ for the sun's mass. The current theory on the other hand gives an upper limit to ρ only in the smaller stars; above a certain mass the curves no longer cross, and it would seem that as the star's energy supply gives out, it must go on contracting to ever higher density—until the space becomes so much curved that the terms "contraction" and "density" lose all meaning.

† *Monthly Notices, R.A.S.* **91**, 444 (1931). We may take the density ρ to be an approximately known multiple of the electron density σ.

CHAPTER XIV

THE COSMICAL PROBLEM

14·1. Waves and Curvature.

Without appealing to the detailed investigations in Chapters XI and XIII, it is evident that the curvature of space-time introduced in relativity theory and the waves of "ψ" introduced in wave mechanics are equivalent. Both devices are used for the same purpose, to represent the distribution of mass and momentum of physical systems. Both are *devices*; it is not suggested that either the curvature or the waves exist in a literal objective sense. The method of waves gives the finer analysis, and is essential if we are treating systems on the atomic scale; but for the mechanics of ordinary macroscopic objects, and for astronomical and cosmical systems, either method is available, and either method should give the same result.

But the result will be expressed in terms of different natural constants. When the relativity method is used, the principal constant involved is the gravitational constant κ. Wave mechanics does not introduce κ, but uses Planck's constant h and the masses m_p, m_e of elementary particles. If we can find *one* problem tractable enough to be solved by both methods, we shall, by comparing the two answers, obtain a relation between the natural constants. I cannot but think that the realisation that a hitherto unrecognised relation exists—that there will be at least one redundant constant when the theories are brought together—is scarcely less important than the ascertainment of its precise numerical form.

In this chapter we shall solve one problem by both methods, and so ascertain the relation between the natural constants. The system which we shall consider is a self-contained static distribution of material particles (protons and electrons) without radiation. By "self-contained" we mean that it requires nothing external to itself to hold it in equilibrium. In relativity theory the only completely static self-contained system is an *Einstein universe*. In quantum theory a system, which is static and radiationless, must be in its *ground state*. We shall therefore treat an Einstein universe, (*a*) by the ordinary relativity theory, and (*b*) by wave mechanics applied to a system of particles in the ground state.

As in the preceding chapter, we represent matter in equilibrium by standing waves. But there will now be no artificial boundaries, and the waves extend throughout the universe.

In stereographic coordinates the line element of an Einstein universe is†

$$ds^2 = -(1+r^2/4R^2)^{-2}(dx^2+dy^2+dz^2)+dt^2 \qquad (14\cdot11)$$

† Tolman, *Relativity, Thermodynamics and Cosmology*, equations (138·4), (139·3), (139·4).

and the pressure and density are found by the usual method to be

$$\left.\begin{aligned} 8\pi\kappa P &= -R^{-2}+\lambda, \\ 8\pi\kappa\rho &= 3R^{-2}-\lambda. \end{aligned}\right\} \tag{14·12}$$

Let

$$\beta = 1+r^2/4R^2. \tag{14·13}$$

Instead of treating R as a radius of curvature, we treat β as a gauge factor. That is to say, we treat the stereographic projection as the true configuration; the curvature is abolished, and x, y, z are rectangular coordinates in a flat space. The "true" element of length is then $ids_0 = (dx^2+dy^2+dz^2)^{\frac{1}{2}}$, which is β times the measured length ids. Hence the true volume is β^3 times the measured volume, and the true particle density σ_0 is β^{-3} times the measured particle density σ. The measured density of distribution of particles in an Einstein universe is uniform; in the new reckoning this uniform density σ is replaced by $\sigma_0 = \beta^{-3}\sigma$.

Expressing the particle density σ_0 in the flat space as the product of two equal scalar wave functions ψ, ϕ, we have

$$\begin{aligned} \psi = \sigma_0^{\frac{1}{2}} &= \sigma^{\frac{1}{2}}\beta^{-\frac{3}{2}} = \sigma^{\frac{1}{2}}(1+r^2/4R^2)^{-\frac{3}{2}} \\ &= \sigma^{\frac{1}{2}}\{1-3(x^2+y^2+z^2)/8R^2+\ldots\}. \end{aligned}$$

From this we calculate the eigenvalue of ∇^2 at the origin

$$\nabla^2 = -9/4R^2 \tag{14·141}$$

so that, by (13·24),

$$\mathbf{P} = 3/4R^2m. \tag{14·142}$$

To obtain the expectation value P, we must multiply the eigenvalue (at the origin) by the probability that the entity represented by ψ, ϕ is in unit volume at the origin. The whole volume of space in natural measure is

$$V = 2\pi^2R^3. \tag{14·143}$$

Since the distribution is uniform, the probability associated with any unit natural volume is $1/V$. At the origin $\beta = 1$, and the two reckonings of volume agree. Hence

$$P = 3/4R^2mV, \quad \rho = 9/4R^2mV, \tag{14·144}$$

by (13·35).

Since we have represented the whole density σ_0 by one pair of wave functions, we have treated it as the probability distribution of a single scalar particle. The formulae may be regarded as referring to an ideal one-particle universe; or, if applied to the actual universe, m is the mass-constant of the equivalent single particle. The extension to a system consisting of N particles which obey the exclusion principle will be treated in § 14·2.

In Chapter XIII only part of the energy was represented by standing waves, the rest energy being taken care of by the external wave function. But here the whole curvature of the Einstein universe has been replaced by

wave functions ψ, ϕ in a flat space-time. Thus the density $\rho = 3P$ in (14·144) is the whole of the density.

Setting $\rho = 3P$ in (14·12) we obtain

$$\lambda = \tfrac{3}{2}R^{-2}, \quad 8\pi\kappa\rho = \tfrac{3}{2}R^{-2}. \tag{14·151}$$

Let us now revert to the usual standpoint, and regard R as a radius of curvature. The waves are then abolished; they were a representation of the projection factor β which is no longer used. Since there are no standing waves, $P = 0$. Inserting $P = 0$ in (14·12) we obtain

$$\lambda' = R^{-2}, \quad 8\pi\kappa\rho' = 2R^{-2}, \tag{14·152}$$

accents being used to distinguish this from the previous reckoning. We have

$$\rho' = \tfrac{4}{3}\rho. \tag{14·153}$$

The change of reckoning from ρ to ρ' depends on the change of the arbitrary constant in the energy tensor from λ to λ'. We have seen (§ 13·1) that this signifies a change of zero point from which energy and pressure are reckoned. When we represent the curvature by waves we introduce a pressure, uniform throughout the universe, which (according to our usual outlook) is fictitious. The change of zero point gets rid of this pressure, and at the same time changes ρ to ρ'.

The pressure (in ordinary reckoning) is not necessarily zero in an Einstein universe. Static universes with non-vanishing pressure are obtained by taking λ slightly greater than R^{-2}. But we are here treating a radiationless universe, which is accordingly at zero temperature, so that the pressure is zero.†

By (14·144) and (14·153), $\rho' = 3/R^2mV$. The total mass or energy of the Einstein universe in ordinary reckoning is therefore

$$M' = \rho' V = 3/R^2 m. \tag{14·161}$$

There is further the well-known relation, obtained from (14·152) and (14·143),

$$\kappa M' = \tfrac{1}{2}\pi R.$$

Thus if M' or R is given we can determine the mass-constant m, which occurs in the definition of the energy operator.

If there are two reckonings of m as there are of ρ and M, viz. $m' = \frac{4}{3}m$, (14·161) can also be written

$$M' = 4/R^2 m'. \tag{14·162}$$

The origin of the factor $\frac{4}{3}$ is that, since we are treating static conditions, the gauge transformation is applied in three dimensions only. The gauge transformation in § 13·7, which gave the most straightforward formal

† According to (13·27) the pressure at zero temperature is not strictly zero. For the Einstein universe this residual (degeneracy) pressure is found to be of order $10^{-26}\rho$, and could be neglected. It is, however, the formula (13·27) which should be corrected by this small amount (to allow for the cosmical curvature); and the degeneracy pressure in the Einstein universe is strictly zero.

connection between waves and curvature, was in four dimensions. Consequently the four-dimensional volume element is changed in the ratio β^3 in this section, whereas it was changed in the ratio β^4 in § 13·7. Another example of this factor occurs in the two modes of reckoning magnetic energy (12·85) and (12·86), according as the field is strong enough to render it necessary to treat the problem in four dimensions or weak enough to be treated as a perturbing effect in a three-dimensional problem.

14·2. Analysis into Particles.

In the foregoing calculation we have treated the uniform density ρ of an Einstein universe as though it were the probability distribution of a single particle of mass-constant m. We shall now suppose that the space of radius R is occupied by N' elementary scalar particles. By the exclusion principle these will be represented by orthogonal wave functions, each particle having a separate wave function. To obtain the ground state of the distribution we must select the wave functions of lowest energy.

The wave functions can be classified similarly to those of an atom, except that we are not troubled by the duplication of states due to opposite directions of spin. (That is taken care of automatically when the N' scalar particles are replaced by $4N'$ elementary vector particles.) In the atom the waves are concentrated into small volume by a controlling Coulomb field. Here the gauge factor β plays the part of a concentrating force; or, what comes to the same thing, the waves spread to the natural limit imposed by the finitude of space. In the stereographic projection of the uniform sphere we have a concentration of density towards the origin and a thinning out at great distances, just as in the atom; we may look on it as an illusion of projection, but mathematically the effect is the same as if the distribution were being held together by a central force.

The lowest state (K state) is the projection of a uniform spherical distribution; and the investigation of § 14·1 is directly applicable to it. We need not treat in detail the successively higher states, which will be projections of spherical harmonics in four dimensions. We proceed at once to the highest of the N' occupied states, which we shall call the *limit state*. We denote the energy of the limit state by m_2. For the actual universe the limit state corresponds to very high quantum number (about 10^{26}), so that the energy levels have there become practically continuous. For the ordinary problems of physics (excluding cosmical problems) there is an inexhaustible supply of particles with energies practically equal to the limit energy m_2.

The mean energy of a particle is

$$\overline{m} = \tfrac{3}{5} m_2. \qquad (14·21)$$

This is a general formula applying to all systems of orthogonal functions in three dimensions. For the rectangular waves, treated in the box problem,

it is contained in the formula $\overline{\varpi}^2 = \frac{3}{5}r^2$ (13·26); it there follows from the fact that the elementary wave functions form a lattice of uniform density in ϖ-space. It is well known that the density of the lattice remains invariant when other orthogonal functions are used instead of Fourier functions, and when the controlling field of potential is varied; or, as it is usually expressed, each eigenfunction occupies a cell of phase space of volume h^3. The only addition we need make here is that, since $\rho = 3P$ for standing waves, the ratio $\frac{3}{5}$ between the mean contribution and the limit contribution to the pressure applies also to the energy density.

In treating the energy levels in an atom, recourse is had to a device known as the *self-consistent field*. The electrons are put one by one, not into the field of the nucleus alone, but into the field provided by the averaged distribution of the whole system. We use the same principle in analysing the universe into particles. The whole distribution determines the radius R of the space, into which we put the particles one by one; and the controlling field β, which depends on R, is likewise determined by the complete distribution.

Let us glance at the alternative of building up the Einstein universe synthetically. The first particle takes up a distribution of self-equilibrium—a miniature Einstein universe consisting of one particle. We can calculate its energy by (14·162), since in this case $M' = m'$. But the calculation is of little value, since the K state in the completed system will have an altogether different energy. Each particle that is added, not only contributes its own energy, but modifies the energies of all the preceding particles.

By the use of the self-consistent field the energies of the particles are defined in such a way that they are precisely additive. This is important in connection with the question, whether the whole density of the distribution is represented by $\Sigma \psi_k \phi_k$ or by $(\Sigma \psi_k)(\Sigma \phi_k)$, ψ_k and ϕ_k being elementary wave functions. In ordinary applications, such as the problem of the rectangular block, it does not matter; because, owing to the orthogonality of the wave functions, the cross terms yield zero integrals over the small volume that is being considered. But when the waves extend over the whole universe, the vanishing of the integral over the whole universe is not a sufficient reason for neglecting the cross terms locally. When we use the self-consistent field the density is $\Sigma \psi_k \phi_k$. For the wave functions, and the energy derived from them, are calculated on the supposition that all the particles are present and producing the field β. The cross-energy with all the other particles, as well as the self-energy of the kth particle, is already incorporated in $\psi_k \phi_k$; and we do not require additional terms $\psi_k \phi_{k'}$ to represent it.

This is the same outlook as in § 12·6, where the mutual energy invariant of two systems is replaced by self-energy invariants. In fact the energy attributed to a particle must necessarily originate as a mutual energy of some kind. It is only when the distribution of the other particles differs from the

standard distribution used in calculating the supposed self-energy, that the mutual energy terms have to be recognised explicitly.†

A precaution is necessary is using the self-consistent field, if we are to avoid counting the energy twice over. In a field of fixed potential ϕ the energy of a particle is reckoned as $m\phi$; but in a field produced by other particles the energy is reckoned as $\frac{1}{2}m\phi$, the other half being allotted to the particles producing the field.

We may summarise the progress of ideas as follows. The mechanical properties of macroscopic matter forming an Einstein universe are usually represented by curvature of space-time; but we can, if we prefer, analyse it into particles and embody the mechanical properties in wave functions. Energy and pressure represented by wave functions must be omitted from the curvature (§ 11·7); thus the complete wave representation consists of wave functions in flat space-time. We therefore consider an alternative representation in flat space-time obtained by stereographic projection of the distribution. In this representation there is a concentration towards the origin, as though a controlling force prevented the matter from spreading away to infinity. We analyse this density distribution into orthogonal wave functions, representing particles which obey the exclusion principle. All the particles must be fully present in the distribution, up to a limiting energy determined by the number of particles; otherwise there would be a conflict between the relativity conditions and the wave mechanical conditions for a static or ground state. Conceptually each wave function or state could be occupied or not; but actually the decision that the states shall all be occupied was taken at the beginning when we chose a radiationless static universe for projection. Thus the disturbance of a particle by the particles occupying other states has been taken into account from the beginning. The test of consistency is that the sum of the densities corresponding to each state shall reproduce the density originally postulated.

14·3. Threshold Energy.

In elementary quantum theory the system under discussion is always treated as an independent addition to the rest of the universe. If the quantum physicist ever remembers that there is a "rest of the universe", he treats it as an ideal background which will not interfere with his system. He does not picture it as consisting of N' other particles competing with his own particles for the states of lowest energy; nor does he contemplate the possibility of his own particles dropping into vacant levels in the background. The background is treated as impermeable.

This is equivalent to assuming that the "rest of the universe" is in the

† Thus perturbation theory will involve $(\Sigma\psi_k)(\Sigma\phi_k)$ since it deals with deviations from the distribution which furnishes the adopted self-consistent field.

ground state; so that there are no excited particles to drop into the vacant levels in the added system, and no vacancies for the added particles to fall into.

When another (scalar) particle is added to the system of N' particles in the ground state, the lowest vacant energy level is at the limit energy m_2. The added particle must be endowed with an energy m_2 in addition to any transferable energy it may possess. Thus m_2, which is the *limit energy* of the background particles, is the *threshold energy* of the added system. There is thus no interference between the added particles and the rest of the universe. The energy levels in the added system begin at m_2; the particles forming the rest of the universe are all placed below this threshold level.

The threshold energy m_2 is the rest energy or proper mass of an added particle. This applies to scalar particles; it is modified in accordance with the theory of Chapter XII when we substitute charged particles. The added particle may be excited above the threshold level, and so possess additional (kinetic) energy. The transferable energy of a particle is really its energy of excitation in the universe-atom; though it is, of course, not usually regarded that way.

Although the elementary equations of quantum theory postulate that the particles forming the rest of the universe are in the ground state, that does not mean that the theory is only applicable if the universe is an Einstein universe. Of necessity an elementary equation refers to idealised conditions. The ground state is the fixed standard background; and any deviation from the ground state must be explicitly described as an *addition* to the fixed background, and as such taken account of as part of the added system. Usually these additions are called "fields".

For example, let the added system consist of particles in thermodynamical equilibrium at temperature T. Clearly there must be statistical equilibrium between the added particles and the background particles, so that the background particles will be excited. But the excited background is not treated as such. It is described as a fixed impermeable background plus a field of radiation. Exchanges of energy between the added particles and the excited background are described as exchanges between the added particles and a field of radiation.

Radiation is the name under which vacancies in the sub-threshold levels are taken into account in current theory. I can see no possible doubt about this identification. For in problems of this type radiation is the only entity, besides the particles, that is mentioned as an addition to the fixed background. So that, if the vacancies are not taken into account as radiation, they are not taken into account at all—which would clearly be a gross error rendering agreement with observation impossible.

I think that this is the natural approach to a quantum-relativistic theory of radiation. But as radiation is rather apart from the main subject of these

investigations, I have not found time to develop the theory. I do not suppose that progress will be easy; for it is a long step from radiation in an Einstein universe to radiation in the concentrated systems where it is studied in practice. That radiation is the positive aspect of the vacancies caused by excitation of the unspecified (scalar) particles seems to me indubitable; but what I now add is an unchecked first impression.

Presumably a single vacant level constitutes a *photon*. The energy of the photon is that which is released when the photon is abolished by a limit particle falling into the vacancy, i.e. the depth of the vacant level below the limit energy. Since the energy levels are practically continuous near m_2, a photon can have practically any energy up to m_2. I doubt if m_2 is a genuine upper limit; further investigation would be required to ascertain whether it still applies in less idealised conditions.

In non-equilibrium conditions the added particles may be localised as *wave packets*; similarly the vacancies may be localised as *wave pockets*. The wave pocket seems to be nothing more than the partial localisation of a photon, manifested, for example, in the observation of individual X-ray effects in an expansion chamber.

The particles which should have occupied the vacant levels will exist at or above the limit level. (A vacancy, in the sense of annihilation of a particle, is not a photon but an impossibility.) If a photon forming part of a field of radiation is absorbed by an atom, we can picture it as continuing its existence in the same form inside the atom, namely, as a vacancy at one of the lower levels in the atom. In such exchanges the energy of the photon, measured by the depth of the vacancy below the limit up to which the levels are occupied, does not remain constant, since the ejected particle may carry off some of the energy.

Leaving the subject of radiation,† we return to the added particles. The added system is formed by specifying certain particles either by wave functions or macroscopically. The particles to be specified must always be taken from the uppermost level, so as to leave no hole in the background. Since the levels near m_2 are exceedingly close together, there is for ordinary purposes a practically unlimited supply of particles of rest mass m_2; and the specified particles will have this proper mass. But when we extend the formulae to cosmical systems, we must ultimately come to the deeper strata. Finally, when we apply our calculations to the whole Einstein universe, and specify the whole of the N' particles (macroscopically), the total rest energy instead of being $N'm_2$ is

$$M = N'\overline{m} = \tfrac{3}{5}N'm_2 \qquad (14{\cdot}31)$$

by (14·21).

† For a deduction that photons (as here identified) obey Einstein-Bose statistics, see § 16·3.

The deficit $\frac{2}{5}N'm_2$ is the (negative) *gravitational potential energy.*

Gravitational energy is therefore the exclusion effect treated in § 14·2, but viewed from the highest level instead of from the zero level. Remembering that the unspecified particles have to form an impermeable system, each successive particle that is transferred to the specified system must be taken from a slightly deeper level. The rest energy of n particles is slightly less than n times the rest energy of one particle. When we come across this phenomenon observationally we attribute it to gravitational energy. We have here considered only the simple case of an Einstein universe, and have not entered into the complications due to concentrating the specified particles. That is scarcely necessary; for we are not putting forward a new "explanation" of gravitational energy. The relativity representation of a gravitational field as curvature of space-time was the starting point of the whole investigation. What we have now reached is the same phenomenon viewed from the topsy-turvy outlook of elementary quantum theory.

Exclusion energy, interchange energy, potential energy are different ways of regarding the same thing. Up to the present we have been concerned only with gravitational potential energy; but in Chapter XV we shall find that electrical potential energy has the same origin.

There is reason to think that our expanding universe is rather far removed from the Einstein state and that a de Sitter universe would be a better approximation. It may be suggested that this, whilst not invalidating the use of an Einstein universe as a standard background, may cause it to be inconvenient in practice. But the local conditions of our experiments always differ much more widely from those of an Einstein or a de Sitter universe than these do from one another. The density in an Einstein universe is $3{\cdot}32.10^{-27}$ gm. cm.$^{-3}$ The use of an Einstein universe as the standard condition, to which the exact equations refer, means that gravitation corresponding to this density has been allowed for in the equations. The actual field in any practical problem is always very much greater, and it is unimportant whether the standard field is deducted or not.

14·4. Positrons and Negatrons.

The scalar wave functions, which we have been considering, can each be analysed into four vector wave functions representing protons and electrons. We may therefore have at any level a vacancy due to the absence of a proton or electron, instead of the absence of the whole scalar particle. We have seen that the absence of a scalar particle is represented as the addition of a photon. According to Dirac the absence of an electron is represented as the addition of a *positron.*

I think that the present investigation throws new light on the meaning of the vast number of occupied negative energy levels, which Dirac postu-

lated in his theory of the positron. The levels are described as negative, because energy is now reckoned from the threshold energy as zero.† This is in accordance with the outlook of elementary quantum theory described in § 14·3; the systems which it treats are additions to the fixed background, and the zero of energy is therefore transferred to the threshold level at which the additions start. Of these negative energy states, Dirac says‡

> These assumptions require there to be a distribution of electrons of infinite density everywhere in the world. A perfect vacuum is a region where all the states of positive energy are unoccupied and all those of negative energy are occupied....The infinite distribution of negative-energy electrons does not contribute to the electric field.

We have reached an altogether different view of the way in which the negative energy levels are filled. They are occupied by the large amount of matter of the universe not specifically mentioned in the equations which Dirac was developing and applying. This matter is ignored, i.e. treated as a fixed background, in the idealised equations; and such treatment automatically relegates it to the levels below the threshold which it must fill completely. Of course, in the actual universe this is far from true; but the idealised equations are a far from complete representation of the conditions, and must be supplemented by terms representing fields of radiation, etc.

The chief points on which we go beyond Dirac's theory of the positron are:

(1) Our negative energy levels are occupied by protons and electrons equally. Dirac supposes them to be occupied by electrons only, with the result that there is an infinite negative charge to be suppressed in an arbitrary way.

(2) Our theory places protons and electrons on the same footing, and therefore definitely predicts the existence of *negatrons* or negative protons, unless there is some unforeseen limit ($< 2m_p$) to the amount of energy that can be expended in a single process of excitation.

(3) The number of negative energy levels is not infinite; it is known precisely (§§ 14·7, 16·8).

When an electron annihilates a positron, the sum of their energies ($\geqslant 2m_e$) is emitted as radiation. In Dirac's theory it is assumed that this is equal to the difference of energy level, as though the falling of the electron into the vacant level were the same kind of transition as that which gives a spectral line of an atom. But in the annihilation of an electron and positron in free space, *two* photons are emitted. The process

† This, however, is not the negative energy to which Dirac refers. A positron has rest energy m_e, and he therefore supposes that the particle whose absence it represents would have had rest energy $-m_e$. As explained at the end of this section, we do not accept this conclusion.

‡ *Quantum Mechanics*, 2nd ed., p. 271.

is therefore not comparable with an ordinary transition in which the whole energy is emitted as one photon.

In our theory the depth of the vacant level below the zero level gives only the excess of the energy of the positron above its rest energy m_e. A vacancy just below the limit level corresponds to a positron almost at rest. We have seen (§ 6·7) that owing to the idempotency of the stream vector of an elementary particle, the vector J representing charge, current, etc., is often confused with the vector J^2 representing energy, momentum, etc. The absence of an electron whose stream vector would have been J is equivalent to the presence of a positron with stream vector $-J$; the reversal of the charge and current is thus duly indicated. But the energy and momentum of the electron and positron are J^2 and $(-J)^2$, so that they are the same; in particular, the quarterspur which represents the proper mass is the same. Our later developments of the theory of the origin of mass have thrown some more light on this point. We recall that energy is furnished to a system by specifying it (§ 11·6), and that the energies m_p, m_e are additions made at one stage of the specification, viz. when a neutral particle is specified as a particle with definite charge and spin. So far as this addition is concerned, it makes no difference whether the original neutral particle had a positive or a negative existence.

14·5. Masses of the Added Particles.

It is useful to distinguish two systems of application of wave mechanics:

A. Cosmical System. This treats the specified and unspecified particles as one great system of N' particles. Owing to the exclusion principle the particles occupy various energy levels; and in the ground state their energies extend up to a limit m_2.

B. Local System. The zero of energy is moved up to the level m_2, and the local system is an addition above that level. On the negative side of the new zero there remains the completely filled set of levels (now negative) which is never disturbed. This is variously regarded as a comparison fluid, an impermeable background, a pure inertial field, or a static spherical space, on or in which to erect system B.

A disturbance of system A from its ground state necessarily involves defect below the limit level as well as excess above it. But nevertheless in system B the levels are always assumed to be completely filled; and the defect as well as the excess is represented as an addition above the new zero. Suitable entities, such as radiation, positrons, electromagnetic potential, are introduced to represent the defects when considered as additions. In Newtonian theory an irregular gravitational field is also regarded as an addition to the fixed background. If we follow Einstein's theory, it is not an addition to but an *abandonment* of the fixed background. A space of variable,

and in general non-static, metric is substituted for the uniform static background. System B is then inapplicable and it is necessary to revert to system A.

We shall now consider more closely how the electrons and protons of the local system are created as an addition to the fixed background. We take a neutral particle (a quarter of a scalar particle of system A) at the limit energy, and "specify" it. That is to say, we assign to it wave functions giving its probability distribution as modified by incorporating observational information as to its charge, spin, momentum, position, etc. As a neutral particle it was equally likely to be a proton or an electron; we now specify it definitely as an electron, say. This involves the use of vector wave functions. The vector wave functions also make the particle mobile in the ordinary sense, i.e. characterised by a stream vector to which Lorentz transformations are applicable. The scalar wave functions can only express mobility of the type represented by standing waves.

Let us now regard spherical space as the frame, the original probability distribution of the neutral particle as the partial comparison fluid, and the specified distribution as the object particle. The observable relations are contained in the double wave vectors Ψ, X of the combined system. It is in these that the observational information is directly incorporated; for example, our observations create wave packets in Ψ, X. Current wave mechanics replaces these by simple wave vectors ψ, χ for the object particle and scalars ϕ, ω for the comparison fluid. As shown in Chapter XII this imposes the relation

$$10m^2 - 136mm_0 + m_0^2 = 0 \qquad (14{\cdot}51)$$

between the mass m of a proton or electron in the local system and the mass m_0 of the neutral particle.

Remembering that the addition of two systems is represented by multiplication of their wave functions, we notice that whereas Ψ describes a modification of the distribution, ψ describes what has been added by the modification. The reduction from Ψ to ψ is really a casting out of the original background particle. If we define an electron or proton to be the entity represented by ψ, it is a pure addition to the background. Accordingly the simple wave vectors are the appropriate representation in system B. This may be more easily seen if we use energy invariants as in § 12·6. By specifying with double wave functions a neutral particle of system A which had an energy invariant m_0^2, we obtain a particle, also in system A, with energy invariant $136mm_0$. This is depicted in system B as having an energy invariant $10m^2$. By (14·51) there is left an amount m_0^2 to provide for the original particle which remains in the background. Thus the procedure of specification, besides providing a particle in system B, stops up the hole in the background.

Our result that the rest mass of a particle is to be identified with the threshold energy m_2 applies to scalar particles. The mass becomes multiplied by $\frac{1}{2}(m_p+m_e)/m_0 = 136/20$ when the scalar particles are replaced, as they eventually must be, by protons and electrons. We may regard the factor 136/20 as a transformation of the scale of reckoning in passing from system A to system B.

We can now see the complete connection between the masses of the elementary particles in quantum theory and the conception of mass as curvature in relativity theory. In quantum theory mass corresponds to the periodicity of the waves. The only direct connection of periodicity and curvature is the periodicity which arises from "going round the world", the corresponding wave length being $2\pi R$. The corresponding mass in C.G.S. units is $h/2\pi Rc$. At first sight this seems far too small to be important. But we have to remember that there can be only one scalar particle to which this applies; and the exclusion principle forces succeeding particles to have periodicities corresponding to the higher harmonics. The result is that an average particle has a mass comparable with that of an electron or proton. But there is still the factor 136/20 to be applied before we reach the masses ordinarily recognised. It is not necessary to go over again the explanation of its occurrence in Chapter XII. We may note, however, that since the rest mass is not an intrinsic attribute of the particle but represents the energy of the particle in an assembly in statistical equilibrium, it depends on the number of degrees of freedom which share in the equipartition of energy. The factor is really a compensation for adopting (in current theory) a simplified representation in which the recognised number of degrees of freedom is greatly reduced by the substitution of simple for double wave functions.

14·6. The Standard Mass m_0.

It is perhaps necessary to remind ourselves that we are not putting forward a new theory of natural phenomena; we are comparing two mathematical methods of treating the same system. Primarily we are treating an Einstein universe, i.e. a uniform static distribution of matter extending indefinitely; but in order to compare our scale of measurement with that used in quantum theory we insert in it a microscopic "added system". From the ordinary point of view (system B) the added system is the centre of attention; the equations used to describe it in elementary quantum theory take it for granted that its surroundings are uniform and static, as here supposed. The surroundings are normally represented by a metrical field $g_{\mu\nu}$, which in the uniform conditions postulated corresponds to a spherical space of radius R. The value of R is involved because, by the macroscopic theory, the material standard of length used in actual experiments on the added system takes up

an extension which is in a definite ratio to the radius of curvature of the corresponding three-dimensional section of space-time.† Thus we have to treat an added system represented by waves in surroundings represented by curvature. We now introduce a mathematical transformation of the problem, which represents them both together by waves (system A). We first represent the uniform surroundings, or Einstein universe, by waves. The added system is then formed primarily by a modification of these waves—a specification by double wave functions. But the modification can also be represented by the original waves together with additional waves; and the additional waves form the usual description of the added system.

It is important to realise that the physical system studied in this chapter is precisely the physical system which a quantum physicist treats at the beginning of his subject. We are treating an Einstein universe; that is all—except that we forgot to mention that one of its particles must be excited (one will be enough) in order that the natural constants h, m_e, etc. may appear in the problem. The quantum physicist takes as his most elementary problem one particle in free space; that is all—except that he forgot to mention that there are some 10^{79} other particles present in an equilibrium state. (For he believes that his formulae can be experimentally verified without destroying all the particles in the universe except one.) Our problems are identical; but our respective forms of absent-mindedness show that we view them from a different outlook. The two outlooks correspond to system A and system B.

We must notice one point in the transformation which will be important in the numerical calculations about to be made. The limit state of system A becomes the K state of system B. We define the K state as that in which a particle has uniform probability distribution over spherical space. The threshold energy m_2 is that of a particle definitely at rest in system B and therefore having entirely uncertain position. The added particles in system B do not exclude one another, except for the slight negative exclusion effect represented by gravitational potential energy; thus, except in so far as the position is specified, any reasonable number of them can occupy the K state of uniform distribution. This change of K state occurs when we pass from the universe with $\rho = 3P$ to the universe with $P = 0$—the change which introduces the factor $\frac{4}{3}$ (14·153). In abolishing the pressure, the energy previously represented by standing waves is replaced by curvature. We must not duplicate the representation by waves and curvature. Energy up to the threshold energy being represented by curvature, only the excess remains to be represented by waves. Thus in system B a particle at rest is represented by a uniform distribution of probability, instead of by standing waves corresponding to a spherical harmonic of high order.

† *Mathematical Theory of Relativity*, § 66.

The connecting link between systems A and B is the mass m_0 of the neutral particle. In system B it has been defined as the mass belonging to a simple (progressive) wave function which possesses only an algebraic phase; equation (14·51) was derived as the necessary relation between the mass m_0 of such a wave function and the mass m ($=m_p$ or m_e) of a vector wave function both referred to the geometrical frame. In system A our elementary scalar wave functions have mass m_2; but the waves are of a somewhat different character, being standing waves in three dimensions instead of progressive waves in one dimension. It would not have been possible to amalgamate *progressive* waves in three dimensions without introducing matrix phases.

We already know that the scalar particle m_2 corresponds to four neutral particles m_0, which on specification yield two protons and two electrons. There is an apparent discrepancy between the *four* particles and the *three* dimensions of the standing waves. This is rectified by the factor $\frac{4}{3}$ found in (14·153). It is true that $m_0=\frac{1}{3}m_2$; but in ordinary reckoning m_2 is replaced by $m_2'=\frac{4}{3}m_2$, so that

$$m_0=\tfrac{1}{4}m_2'. \tag{14·61}$$

It is convenient to get rid of the factor $\frac{4}{3}$ by introducing an intermediate system A', which is the same as system A except that all masses are multiplied by $\frac{4}{3}$.

Similarly we introduce an intermediate system B' consisting of neutral particles with one-phase wave functions. The multiplication of the mass by 136/20 then occurs in passing from system B' to system B.

We have then four systems defined as follows. System A is the direct result of the analysis of the curvature into elementary scalar wave functions constituting standing waves in three dimensions. In system A' we convert the Einstein universe with $\rho=3P$ into an Einstein universe with $P=0$, thereby increasing all energies in the ratio $\frac{4}{3}$; this increase is the same as if the energy had been calculated for standing waves in four dimensions instead of three, and we shall regard it in that way. In system B' some of the limit particles of system A' with energy m_2' are taken as the basis of an added system; the scalar particle with (effectively) four degrees of freedom is for this purpose divided into four neutral particles with one-phase wave functions and energy $m_0=\frac{1}{4}m_2'$. In system B the ordinary vector wave functions are introduced and the total mass becomes multiplied by 136/20.

In system B' the coefficient m in the energy operator $-m^{-1}\delta^2/\delta x_\beta \delta x_\alpha$ can be identified with m_0. This follows from the elementary formula for the plane waves of a neutral particle

$$m_0=-id/ds=-m_0^{-1}d^2/ds^2.$$

This must also apply to system A'. But in system A the unit of mass is

altered so that $m=\frac{3}{4}m_0$. Thus (14·144), which refers to system A, becomes

$$P=1/R^2m_0V, \quad \rho=3/R^2m_0V. \tag{14·621}$$

Hence the corresponding total mass in system A' is

$$M'=\tfrac{4}{3}M=\tfrac{4}{3}\rho V=4/R^2m_0, \tag{14·622}$$

which is the more explicit form of (14·162). This is the expression for the energy of a scalar particle forming a uniform distribution in spherical space of radius R, i.e. it is the energy of the K state. We shall therefore change the notation M' to m_K. It is convenient at this stage to insert the factor h/π required to reduce masses to C.G.S. units.† We therefore write (14·622) as

$$m_K=\frac{4}{m_0R^2}\left(\frac{h}{\pi}\right)^2. \tag{14·63}$$

As we shall not further use system A, we shall omit the accents previously used to distinguish quantities referred to system A'.

Consider first a universe which consists of one scalar particle. There can be no absolute comparison of standards in different universes; we shall therefore arbitrarily define the corresponding units to be such that m_0 and R are the same as in the actual universe. The masses m_2, m_p, m_e which are numerically connected with m_0 will also be the same. But the natural constants κ, h may be different. We denote their values in the one-particle universe by κ_1, h_1.

In the one-particle universe, the K energy, the limit energy, and the total energy coincide, so that $m_K=m_2=M$. Hence, by (14·63),

$$m_2=4m_0=\frac{4}{m_0R^2}\left(\frac{h_1}{\pi}\right)^2, \tag{14·641}$$

so that

$$m_0=h_1/\pi R. \tag{14·642}$$

Returning to the actual universe, let

$$p=\tfrac{3}{5}N'.$$

Then by (14·31) $m_2=M/p$. That is to say, the wave function representing a single scalar particle uniformly distributed over the sphere‡ will have only $1/p$th of the energy required to produce the curvature $1/R$. In (14·641) a single particle produced the curvature $1/R$, and the coefficient m^{-1} of the energy operator was then $m_0^{-1}(h_1/\pi)^2$; we must therefore take the coefficient to be $1/p$th of this, namely $m_0^{-1}(h/\pi)^2$, where

$$h^2=p^{-1}h_1^2. \tag{14·65}$$

From (14·642) and (14·65)

$$m_0=h\sqrt{p}/\pi R=h\sqrt{\tfrac{3}{5}N'}/\pi R. \tag{14·66}$$

† We still omit the constant c. For the factor h/π, see § 9·6.

‡ As explained earlier in this section the threshold energy m_2 corresponds to uniform probability distribution over the sphere in system B. System B corresponds to the ordinary equations of elementary quantum theory which define the constant h.

We may note also that $\kappa_1 = p\kappa$, since the well-known condition for an Einstein universe, applied firstly to the one-particle and secondly to the N'-particle universe, gives

$$\kappa_1 m_2 = \tfrac{1}{2}\pi R = \kappa p m_2. \tag{14·67}$$

Thus κ/h^2 is independent of N'.

14·7. Numerical Solution.

Since each scalar particle corresponds to four elementary particles, the total number of protons and electrons is $N = 4N'$. Inserting the constant c to reduce to C.G.S. units, (14·66) becomes

$$m_0 = h\sqrt{\tfrac{3}{5}N}/2\pi Rc. \tag{14·71}$$

We have also, by the ordinary relativity formula for an Einstein universe,

$$\tfrac{1}{2}\pi R = \kappa M/c^2 = \tfrac{1}{2}\kappa N(m_p + m_e)/c^2 \tag{14·72}$$

together with

$$m_p + m_e = (136/10)\, m_0 \tag{14·73}$$

by (14·51).

From (14·71) and (14·72)

$$\frac{\sqrt{\frac{3}{5}N}}{R} = \frac{2\pi c m_0}{h}, \quad \frac{N}{R} = \frac{\pi c^2}{\kappa(m_p + m_e)},$$

whence

$$\sqrt{\tfrac{5}{3}N} = \frac{hc}{2\kappa m_0 (m_p + m_e)} = \frac{136}{10}\,\frac{hc}{2\kappa(m_p + m_e)^2} \tag{14·74}$$

by (14·73). We shall find in Chapter XV that $hc/2\pi e^2 = 137$; so that the result can also be written as

$$\sqrt{\tfrac{5}{3}N} = \frac{136.137}{10}\,\frac{\pi}{\kappa}\,\frac{e^2}{(m_p + m_e)^2}. \tag{14·75}$$

Formula (14·75) is the most suitable for an accurate determination of N from observation. As suggested by W. N. Bond the so-called "observational" values of e/m_e are (from the point of view of wave mechanics) erroneous by a factor $\frac{136}{137}$; but the "observational" values of m_p/m_e are also in error by the same factor, so that the observational values for e/m_p can be accepted as correct. I adopt the values given by W. N. Bond† from recent determinations:

$$e/m_e c = 1{\cdot}7574.10^7 \text{ (uncorrected)}; \quad 1{\cdot}7703.10^7 \text{ (corrected)},$$

so that

$$e/m_p c = 1{\cdot}7703.10^7 \div 1847{\cdot}6 = 9582.$$

For κ the generally accepted value appears to be $6{\cdot}664.10^{-8}$, but it is uncertain to 1 part in 1000. Inserting these values in (14·75), we obtain

$$N = 3{\cdot}1454.10^{79}.$$

Following a suggestion by R. Fürth that the core of the large number N is 2^{256}, we can express the number $\frac{1}{2}N$ of electrons or protons as

$$\tfrac{1}{2}N = 135{\cdot}82.2^{256}.$$

† *Nature*, **135**, 825 (1935).

Since N is an integer by definition, no irrational factors can enter into its composition; and the possible hypotheses as to its constitution are very limited. In view of the close association of the numbers 136 and 256 in the theory of the double wave function, our result makes it probable that the exact value is $\frac{1}{2}N = 136.2^{256}$. In Chapter XVI we shall obtain this number independently from theory.

Conversely, assuming $\frac{1}{2}N = 136.2^{256}$, and using the observational values of e/m_p and c, we can calculate κ. The result obtained by Bond† is

$$\kappa = (6{\cdot}659 \pm {\cdot}0012).10^{-8}.$$

If the ratio of the electrical to the gravitational force between a proton and electron (calculated according to classical theory) is denoted by F, we have

$$F = \frac{e^2}{\kappa m_p m_e} = \frac{136^2}{10} \frac{e^2}{\kappa (m_p + m_e)^2}.$$

Hence (14·75) gives

$$F = \frac{136}{137\pi} \sqrt{\tfrac{5}{3}N}. \tag{14·76}$$

14·8. Alternative Treatment.

If the scalar wave functions ψ, ϕ contain a time factor e^{ikt}, the particle density σ contains a time factor e^{2ikt}. The volume is therefore expanding as e^{-2ikt}, and the linear scale as $e^{-\frac{2}{3}ikt}$. We can therefore connect the representation of mass by waves with the theory of the expanding universe.

Consider an expanding spherical universe whose radius at time t is $R = R_0 e^{f(t)}$. The line element is‡

$$ds^2 = -\frac{e^{2f(t)}}{1 + r^2/4R_0^2}(dx^2 + dy^2 + dz^2) + dt^2, \tag{14·81}$$

x, y, z being the stereographic coordinates at time t. The pressure and density are found to be

$$\left.\begin{aligned} 8\pi\kappa P &= -R^{-2} - 2\ddot{f} - 3\dot{f}^2 + \lambda, \\ 8\pi\kappa\rho &= 3R^{-2} + 3\dot{f}^2 - \lambda. \end{aligned}\right\} \tag{14·82}$$

To represent the above imaginary expansion of linear scale, we must take

$$f(t) = -\tfrac{2}{3}ikt \tag{14·825}$$

so that (14·82) becomes

$$\left.\begin{aligned} 8\pi\kappa P &= -R^{-2} + (\tfrac{4}{3}k^2 + \lambda), \\ 8\pi\kappa\rho &= 3R^{-2} - (\tfrac{4}{3}k^2 + \lambda), \end{aligned}\right\} \tag{14·83}$$

which is a generalisation of (14·12).

† *Nature*, **137**, 317 (1936).

‡ Tolman, *Relativity, Thermodynamics and Cosmology*, equations (150·2), (150·7), (150·8).

As in § 14·1, we replace the curvature by standing waves in flat space, so that $\rho = 3P$. Inserting this condition in (14·83) we have

$$\tfrac{4}{3}k^2 + \lambda = \tfrac{3}{2}R^{-2} = 8\pi\kappa\rho, \tag{14·84}$$

which is a generalisation of (14·151).

The meaning of (14·84) is that, by adopting a slightly lower value of the arbitrary constant λ, we can reserve a portion of $8\pi\kappa\rho$ to be represented by wave functions with a time factor e^{ikt}. The reservation of a portion of the whole energy tensor or matrix for representation by waves has been discussed in § 11·7. Let $1/p$th of the energy tensor be thus reserved, so that

$$\tfrac{4}{3}k^2 = 3/2pR^2. \tag{14·85}$$

There are two ways of viewing (14·84). The macroscopic view is that

$$\lambda = 8\pi\kappa_1\rho, \quad \tfrac{4}{3}k^2 = 8\pi\kappa_2\rho,$$

so that the energy of the wave is provided by a general weakening of the gravitation constant involving a decrease of the gravitational mass of every particle in the distribution. But in that case the factor e^{ikt} belongs to a collective wave function of the whole particle density, and not to an individual particle. The alternative view is that

$$\lambda = 8\pi\kappa\rho_1, \quad \tfrac{4}{3}k^2 = 8\pi\kappa\rho_2.$$

In particular, if $p = \frac{3}{5}N'$, the reserved density ρ_2 amounts to just one scalar particle in the K state. This corresponds to the elementary conception of a single particle represented by waves, with an impermeable background represented macroscopically.

The mass corresponding to a time factor e^{ikt} is

$$m = -(ih/\pi)\,\partial/\partial t = hk/\pi.$$

In the case of a collective wave function, every particle has this amount of mass represented by waves; for its own particle density (equal to the product of its individual wave functions) changes with the expansion of R in the same proportion as the whole particle density. The total mass represented by the waves is therefore hpk/π.† This mass is now replaced by a single reserved particle whose individual wave function must accordingly be given a time factor e^{ipkt}. In ordinary reckoning (system A') the mass of the particle is multiplied by $\frac{4}{3}$, and becomes

$$m' = \tfrac{4}{3}phk/\pi.$$

Hence, by (14·85),

$$m'^2 = \frac{2p}{R^2}\left(\frac{h}{\pi}\right)^2. \tag{14·86}$$

This result must, however, be divided by 2, because the energy density ρ_2 here attributed to the particle is a mutual energy of the particle and the rest

† That is to say, if the individual wave functions are of index 1, the collective wave function is of index $1/p$.

of the system; so that if we sum it for all the particles we obtain twice the total energy. It will be seen from (14·84) that if we reserve in succession the p equal amounts of density ρ_2, the unreserved portion of R^{-2} (which is left to be represented in system A' as curvature) diminishes by equal steps, so that m'^2 decreases regularly from (14·86) to zero. The corrected result (giving the energy reckoned according to the method of the self-consistent field) is therefore

$$m'^2 = \frac{p}{R^2}\left(\frac{h}{\pi}\right)^2. \qquad (14{\cdot}87)$$

The mass m' here found is to be identified with m_0. It is perhaps rather difficult to see that it is m_0 rather than m_2. But it is to be remembered that the mass constant both for the scalar particle and the neutral particle is m_0; and the fourfold energy of the former comes from its additional degrees of freedom. We are here treating the progressive waves in one dimension t; and (14·87) applies to one particle with one degree of freedom. If we represented the four neutral particles by separate wave functions, *each* of them would have the time factor e^{ipkt} and the corresponding energy; just as, when earlier we divided the collective wave function into p individual wave functions, each of them had the energy corresponding to the time factor.

The result (14·87) is accordingly an alternative derivation of (14·66); and the solution in § 14·7 then follows.

If we are not seeking the exact result, it is comparatively easy to obtain a solution of the cosmical problem sufficient to confirm the order of magnitude of the recession of the spiral nebulae.

Consider an object system and N unspecified particles. Observational determinations of the position of the object system must be relative to a physical frame of reference provided by the unspecified particles. Since they are unspecified, they have random positions in the hypersphere of radius R which constitutes space. The mean square deviation of a particle from the centre of the sphere is $\frac{1}{2}R$ in each of the four coordinates; so that the mean square deviation of their centroid is $\frac{1}{2}R/\sqrt{N}$. Thus the physical frame of reference provided by the unspecified particles is such that its origin has a standard deviation $\frac{1}{2}R/\sqrt{N}$ compared with an ideally fixed geometrical frame. By the uncertainty principle the associated standard deviation of momentum is of order $2h\sqrt{N}/R$, or in mass units $2h\sqrt{N}/Rc$.

Equivalently, the motion of each particle is a rotation about the centre of curvature of space and its wave function has a half-quantum $h/2\pi$ of angular momentum in the plane of rotation, or a linear momentum $\pm h/2\pi R$; the resultant of N such momenta in random directions is $h\sqrt{N}/2\pi R$. By either method we deduce a mass m_0 of order $h\sqrt{N}/Rc$ as in (14·71), giving the energy of the physical reference frame referred to the geometrical reference frame. The observed relative motion of a particle referred to the

physical frame is then apportioned between the particle and the frame according to the principles worked out in Chapters XI and XII, and the mass m_p or m_e attributed to the particle is given by the fundamental quadratic (14·51).

It may be objected that by this procedure what is really the same comparison mass m_0 is used over and over again for each particle in the universe. Quite so. That is the secret of the unexpectedly large masses attributed to the particles in quantum theory—unexpected because the constant h which connects mass and curvature in quantum theory is at first sight much too large.

In the exact determination of m_0 in (14·71), several numerical factors, which might easily have been overlooked, have been introduced. We add some further remarks on these.

The factor $\frac{3}{5}$ arises as we have seen from the exclusion energy or gravitational potential energy of the particles, reckoned negative because the maximum exclusion energy or limit energy is our ordinary standard. For the present purposes N' excluding particles are equivalent to $\frac{3}{5}N'$ non-excluding particles. The only point that arises is whether the same correction should not have been used in (14·72). The answer is that (14·72) is a standard equation of relativity theory, here used to express our results in terms of the observed constant κ. On referring to the derivation of this formula in general relativity theory, we find that M is the sum of the masses of the particles without deduction of gravitational potential energy (which is a non-tensor quantity). As we said at the beginning, our plan has been to solve the problem of a static distribution by two methods; (14·71) is the result of the solution by wave mechanics, and (14·72) the result of the solution by relativity theory.

The factor $\frac{4}{3}$, introduced in passing from system A to system A', is perhaps less troublesome than the others, because it appears analytically and we are not directly concerned with its physical significance. We have seen, however, that it is due to the adaptation of the formulae to static problems in which the time dimension has specialised treatment, whereas the more elementary formulae are based on four coordinates to which Lorentz transformations are applicable. Alternatively we can regard it as due to the condition that, by considering a spherical space of *fixed* radius R, we have suppressed all waves in the radial direction. The actual R of the universe (referred either to an ideal geometrical frame or to a physical comparison standard) is, like all observables, subject to the uncertainty principle, and has a probability spread. But it is treated as fixed because, being the sole linear characteristic of the universe in its ground state, it is the standard by which all other lengths are measured directly or indirectly. This posterior fixity, conferred on R by its adoption as standard, conceals its intrinsic variability.

As soon as the static condition is relaxed, the variability of R (the local radius of curvature) forces itself on our attention; it is duly exhibited in the formulae of general relativity theory. In other dimensions, we pass from the non-static to the static condition by substituting constant standing waves for irregular wave motion. In the R dimension irregular wave motion is simply dropped, and is not replaced by standing waves. I do not suggest that this is an erroneous treatment; but we see how the fourth degree of freedom has come to be suppressed.

To use four-dimensional waves in the present problem would defeat our purpose. The Einstein universe is unstable.† It is in statistical equilibrium only on the understanding that transfer of the energy of the three-dimensional waves into the radial dimension is inhibited. If transfer occurs, we have an expanding or contracting universe; presumably this is a state of pursuit of equipartition of energy—an equipartition which can never be attained, since expansion of scale corresponds to a hyperbolic phase, and the wave functions in the radial dimension are exponentials instead of sines and cosines.

Clearly it is more than a happy coincidence that the energy is multiplied by a factor $\frac{4}{3}$ in passing from system A to system A'. So that before the division of the scalar particle into four elementary particles is actually made, a transformation equivalent to the substitution of four for three dimensions occurs. Perhaps it is better to express it the converse way; the four particles are effectively reduced to three before being packed into the static scalar wave function in three dimensions.

Very difficult questions arise in regard to a factor 2 or $\frac{1}{2}$; and it is only by the closest attention to the physical meaning of the analysis that we can hope to get this factor right. It is complicated at the outset because an erroneous factor 2 occurs in Dirac's theory through a confusion between double-valued and single-valued wave functions. It may be noticed that in the solution of the box problem (13·27) we take the momentum factor to be $ih/2\pi$ whereas for the apparently similar standing waves here used we take it to be ih/π. But in the box problem ψ is an internal wave function describing relative motion—as in the theory of the hydrogen atom; whereas here the scalar wave functions are referred to a geometrical frame. Further subtleties arise in the relation of scalar to neutral particles, in the representation of the particles of system B as an addition to, instead of a modification of, the particles of system A (so that when all the particles are specified we duplicate the universe), and in the liability to count the mutual energy of the particles twice over. The last correction was required in our second derivation of m_0, but not in the first derivation. I can only offer the solution here obtained as the best effort I can make to avoid these pitfalls—with the

† Perhaps the term "metastable" expresses the condition more precisely.

feeling that it will be something of a miracle if I have really escaped them all. It is reassuring that the same value of N is obtained by an entirely independent investigation in Chapter XVI—though even there a question of a factor 2 arises which is not easy to decide.

Considerable light is thrown on these factors by a comparison with the theory of the Stern-Gerlach effect—for which the corresponding factors can be checked observationally (§ 12·8). In particular, the internal wave functions with four dynamical coordinates (comparable therefore with standing waves in four dimensions) are found to have a magnetic energy $\frac{4}{10}$ of that of a corresponding external wave function, whereas the single-phase wave function which defines m_0 would (by the same treatment) have $\frac{1}{10}$ of the energy. The 4 : 1 ratio is shown more directly in the magnetic energy than in the dynamical energy, because the masses assigned to the particles are such as to eliminate these factors and validate the current dynamical formulae which ignore them. The factor $\frac{3}{4}$ shown in (12·86), which corresponds to the transition from strong to weak magnetic fields, is equivalent to that introduced in the present problem in the transition from disturbed to static conditions.

I may add an explanation why it is necessary to treat the universe as composed of N' scalar particles rather than of $4N'$ neutral particles in the main part of the investigation. A neutral particle has an equal probability of being positive or negative; a collection of $4N'$ neutral particles has therefore a probable charge $\pm e\sqrt{4N'}$. This is a large charge which would give a potential throughout the universe of the order 10^4 electrostatic units or 10^6 volts. We cannot trust to probability to provide a neutral universe; we have to build it of units whose charge is definitely zero, i.e. scalar particles.

It is worth noticing that in (14·74) the numerical factor $\frac{1}{2}.\frac{136}{10}$ on the right has a simple relation to the numerical factor $\frac{1}{2}.\frac{10}{3}\,(=\frac{5}{3})$ on the left. The numbers 3, 10, 136 are the numbers of symmetrical components in 2-fold, 4-fold and 16-fold matrices, respectively. I think it might be possible to exhibit the factor on the left as arising in the reduction of Dirac wave vectors to Pauli vectors (which are adequate for describing a static distribution), in the same way that the factor on the right arises in the reduction from double wave vectors to simple wave vectors.

14·9. The Recession of the Nebulae.

Having found N, we can determine the Einstein radius R of the universe by (14·72). The result is

$$R = 1{\cdot}234\,.\,10^{27}\text{ cm.} = 400{\cdot}3\text{ megaparsecs.}$$

The total mass and density in the Einstein state are

$$M = 2{\cdot}61\,.\,10^{55}\text{ gm.} = 1{\cdot}32\,.\,10^{22} \times \text{sun's mass},$$

$$\rho_e = 3{\cdot}32\,.\,10^{-27}\text{ gm.} = 1\text{ hydrogen atom per 500 cu. cm.}$$

The limiting speed of recession of distant objects is $c/R\sqrt{3}$ per unit distance. We obtain

Speed of recession = 432 km. per sec. per megaparsec.

The observational value of the speed of recession of the extra-galactic nebulae (usually given in round numbers as 500 km. per sec. per mp.) is in as close agreement as could be expected.

In terms of the cosmical constant λ the radius and mass of an Einstein universe are $R_e = \lambda^{-\frac{1}{2}}, \quad M_e = \frac{1}{2}\pi\lambda^{-\frac{1}{2}}c^2/\kappa.$

In the mathematical theory of the expanding universe, it has been usual to treat three cases according as the actual mass M is greater than, equal to, or less than M_e. The present theory leads to a different outlook; M is necessarily equal to M_e. If we pay attention to relativity theory only, we have no ground for supposing that any static configuration of the matter of the universe can be found; therefore universes possessing no static configurations, i.e. with $M \neq M_e$, have been considered possible. But when we treat the universe as a collection of N particles obeying the exclusion principle, we see that such a system necessarily possesses a ground state and therefore a static configuration; thus the universes without static configurations are inadmissible.

I suppose that the ground state of the universe has hitherto been pictured as a highly concentrated "atom" with radius negligible compared to the present dimensions of the universe; so that it scarcely came to be considered in connection with the Friedman-Lemaître theory. Our calculation shows that, on the contrary, it has a radius of 400 megaparsecs, and a density equivalent to 2 hydrogen atoms per litre.

We thus discard the naïve idea that in the beginning there was a cosmical constant; and that when the universe was made, the Creator had to decide whether the amount of matter created should be greater than, less than, or equal to the standard mass M_e fixed by it. In the present theory λ is of the nature of a constant of integration adjusted according to the actual mass of the universe.

According to the Friedman-Lemaître theory, if the universe started from the Einstein state of unstable (or metastable) equilibrium, it might either expand or contract. But since the Einstein state is the ground state, contraction seems paradoxical. It would be interesting if a contracting universe could be ruled out in this way. But at present I cannot see that contraction definitely conflicts with wave mechanics; in comparing the ground state of the universe with the ground state of an atom we must bear in mind that the former is unstable and the latter stable. On the other hand, it seems impossible that the contraction should continue indefinitely; and, if it cannot continue, it should somehow be prevented from starting. Formerly

one supposed that the contraction might come to a natural end through quantum complications setting in when the particles became closely packed; but we now see that the "quantum complications" are already in full play in the Einstein universe.

Although the relations between the natural constants have here been calculated for a special distribution of matter, they must hold for the irregular distribution in the actual universe. In determining the constant of gravitation experimentally, the physicist is not forbidden to arrange the matter in his laboratory in any way that suits the experiment; similarly the theoretical physicist is not forbidden to arrange the matter of the universe in any way that makes his calculation easier. In either case the value found for the constant will apply to all distributions, however widely they may differ from those used in the experiment or the calculation. Only we must take note that, in rearranging the matter to suit his purpose, the experimenter cannot, and the theorist must not, violate any law of nature. Thus we have had to defend our rearrangement of the universe as a static configuration, by showing that the matter of the universe is necessarily such that it possesses a static configuration.

These arguments would be complicated if we took note of the fact that, in addition to matter, the actual universe contains a certain amount of radiation. Since the amount is changing, it cannot be regarded as an essential feature of the problem. It can, I think, be eliminated in the permissible "rearrangement" referred to in the last paragraph; but in any case the amount of radiant energy is trivial compared with the whole energy of the universe.

CHAPTER XV

ELECTRIC CHARGE

15·1. Interaction.

In classical physics an interaction between two particles means a difference in the behaviour of one due to the presence of the other. In wave mechanics we study probability distributions, and determine only probable behaviour. Interaction is therefore a difference in the probability of behaviour. The probability distribution of electron A, which specifies its chance of occupying a particular position or possessing a particular momentum, is modified by the presence of electron B.

There are two ways of treating these changes of probability distribution. We represent the actual probability as the product of two factors: (1) the initial probability, or "basis of statistics", and (2) a modifying factor which incorporates any special information supplied. Analytically, the initial probability is the volume of the element of phase space containing the configurations considered; and the modifying factor is given by the product of the wave functions ψ, ϕ. The effect of the presence of electron B can be introduced either in (1) or (2). If we treat the presence of electron B as special information, its interaction is incorporated in the modifying factor. Alternatively we may regard it as a normal circumstance that the electron A, which we are considering, is one of several present in the region; we then incorporate the interaction in the initial probability and adopt a "new statistics" for systems of two or more electrons.

Both alternatives have been commonly employed in describing the interaction of electrons. It is postulated that they repel one another with a Coulomb force. This tends to keep them apart, and thereby modifies the probability distribution which would have been attributed to them as independent systems whose probabilities combine by simple multiplication. It is also postulated that a system of several electrons obeys a new statistics, called Fermi-Dirac statistics.

Both the Coulomb force and the Fermi-Dirac statistics describe an interaction; that is to say, they assign to the electrons a probability distribution of position, momentum and spin different from the distribution for non-interacting particles, whose probabilities are independent and therefore combine by simple multiplication. But the Coulomb interaction is incorporated in the modifying factor, and the Fermi-Dirac interaction is incorporated in the initial probability or basis of statistics. The Coulomb force changes the wave functions, so that they satisfy a modified wave equation containing an extra term called the Coulomb energy. The Fermi-

Dirac interaction gives zero probability to the symmetrical wave functions; these are therefore omitted from the beginning, and the initial probability distribution of momentum is limited to such parts of phase space as correspond to antisymmetrical wave functions. Even if the two interactions were of independent origin it would be desirable to express them in more comparable form. But I think it is obvious that current theory, by treating the interaction of electrons in this piecemeal way, has arbitrarily divided into two compartments a subject which is really one. It cannot seriously be maintained that the Coulomb force, which prevents two slow moving electrons from approaching one another, is an altogether distinct phenomenon from the exclusion principle (contained in Fermi-Dirac statistics) which achieves the same result by forbidding them to occupy the same phase cell.

It is not as though Fermi-Dirac statistics were intended to be a first approximation, giving the probability distribution in the limit when the electrons are so far apart that their Coulomb forces are negligible. The difference between classical and Fermi-Dirac statistics is only important when the electrons are crowded together; and the conditions in which we apply Fermi-Dirac statistics are precisely those in which the Coulomb forces are large. The attitude of current theory is altogether bewildering. It sets up an ideal scheme of statistics only to repudiate it (by introducing a large modification) in the very circumstances for which it is designed.

This separation of the interaction of electrons into two effects strongly resembles the separation of gravitation and inertia in Newtonian mechanics. The latter taught that a body tends to move uniformly in a straight line by its inertia, but is pulled into a different path by the gravitational field. Similarly today quantum physics teaches that electrons tend to take up the probability distribution corresponding to Fermi-Dirac statistics, but are forced into a different distribution by their electrical repulsions. There is need for the same kind of unification of treatment that has proved so successful in the unification of gravitation and inertia.

It is well known that Fermi-Dirac statistics arise from the indistinguishability of the particles concerned. If we are right in believing that Coulomb force is another aspect of the same interaction, it must also arise from the indistinguishability of the particles. To test this we must investigate the precise way in which indistinguishability modifies the enumeration of probabilities; so that we can determine its effect on the wave tensors, and hence on the wave equations which the tensors satisfy. The investigation is carried out in this chapter. We shall find that the effect of the indistinguishability is to introduce an additional term in the wave equation, which turns out to be identical with that which has been adopted empirically to represent the Coulomb energy.

The leading idea in the investigation is that the equations must be invariant for interchange of the indistinguishable particles. Consequently we have a new kind of relativity transformation or "rotation" of the system, bringing about interchange. The interchange can be performed continuously —a gradual transfer of probability from one identification to the opposite identification. The argument of the new transformation is called the *permutation coordinate*, and its conjugate momentum is called the *interchange energy*. On calculating the interchange energy, we find that it agrees with the observational value of the Coulomb energy.

The foregoing argument was the actual starting point of the theory of protons and electrons developed in this book.† It was, I believe, the first introduction of the permutation coordinate and its conjugate interchange energy in wave mechanics. Now that interchange energy is regularly used in practical problems, it is difficult to see why the author's theory of the Coulomb energy of electric charges is still looked upon as a dubious excrescence on wave mechanics. In the equations in current use the identity of interchange energy and Coulomb energy is accepted.‡ I do not understand why an investigation whose results have come to be admitted without question in the formulae in regular use is still commonly alluded to as a "bold speculation".

15·2. Interchange.

The initial difficulty in calculating the effect of interchangeability is that interchange seems to be a discontinuous transformation. But wave mechanics has been successful in replacing quantum "jumps" by continuous analysis, and the same methods are available for treating the jump of interchange.

Denote the coordinates, including suffix coordinates,§ of two particles (not necessarily indistinguishable particles) collectively by x, x'; and let $\Psi(x, x')$ be a wave function of the combined system. Let the operation of interchanging x and x' be denoted by Q, so that

$$\Psi(x', x) = Q\,\Psi(x, x'). \qquad (15{\cdot}21)$$

If Ψ is single-valued, $Q^2 = 1$. But the ordinary relativistic wave function has ambiguous sign (§ 9·6). For when the corresponding space vectors ΨX^* are rotated through 360°, so that all observable characteristics of the system

† "The Charge of an Electron", *Proc. Roy. Soc.* A, **122**, 358 (1929).

‡ Dirac, *Quantum Mechanics*, 2nd ed., p. 228, equation (38). The interchange energy is given as $\frac{1}{2}V_{rs}\{1+(\sigma_r, \sigma_s)\}$, whose eigenvalue is the Coulomb energy V_{rs}. The unitary matrix factor depends on the circumstances of the problem to be treated, and does not affect the identification.

§ When $\Psi_{\alpha\beta}(x_\mu, x_\mu')$ is written as $\Psi(x_\mu, \alpha; x_\mu', \beta)$, we call α, β *suffix coordinates*. Suffix coordinates can take only the four values 1, 2, 3, 4. They are sometimes called spin coordinates; but this name should be reserved for angular coordinates conjugate to spin momenta.

are unchanged, Ψ and X rotate through 180° and become $-\Psi$, $-X$. If the operation of interchange is treated as a rotation, repetition of the operation will bring us to the other branch of the double-valued wave function, so that $Q^2\Psi = -\Psi$. Hence $Q^2 = -1$.

We introduce an "extended wave function" defined by

$$\Psi(x, x', \chi) = e^{\frac{1}{2}Q\chi}\Psi(x, x'). \tag{15·22}$$

Then, since $Q^2 = -1$,

$$\begin{aligned}\Psi(x, x', \chi) &= (\cos\tfrac{1}{2}\chi + Q\sin\tfrac{1}{2}\chi)\Psi(x, x') \\ &= \cos\tfrac{1}{2}\chi.\Psi(x, x') + \sin\tfrac{1}{2}\chi.\Psi(x', x). \end{aligned} \tag{15·23}$$

In particular $\Psi(x, x', \pi) = \Psi(x', x)$; and

$$\Psi(x, x', \chi + \pi) = \Psi(x', x, \chi). \tag{15·24}$$

Thus the particles are interchanged by increasing χ by π. Intermediate values of χ have a simple interpretation. By the usual rule (15·23) represents a superposition of the state $\Psi(x, x')$ with probability $\cos^2\frac{1}{2}\chi$ and the state $\Psi(x', x)$ with probability $\sin^2\frac{1}{2}\chi$. The interchange is therefore represented, not as a sudden jump, but as a gradual transfer of probability from the original to the interchanged state, as χ increases from 0 to π. This is the recognised method of treating transitions between discrete states in wave mechanics.

Suppose that we are describing the distribution of a red particle and a blue particle. We use probability distributions to express our inexact knowledge of their positions. It is appropriate to provide in the same way for inexact knowledge of their colour. To take another example—two golf balls have been driven at a short hole; there is a certain probability that there will be one ball on the green and one in the bunker; there is also an (unequal) probability that the one on the green will be your ball or mine. Therefore, besides stating the probability that there are particles at two points x, x', we can state the probability p that the particle at x is the red one, or the probability $1-p$ that it is the blue one. This is provided for in the extended wave function $\Psi(x, x', \chi)$ which includes a *permutation coordinate* χ such that $\cos^2\frac{1}{2}\chi = p$. In general we treat a state represented, not by one value of χ, but by a probability distribution over the coordinate χ—so that we have a probability distribution of what is itself interpreted as a probability.

The wave function $\Psi(x, x', \chi)$ primarily applies to the general case in which the two particles are distinguished with more or less uncertainty. In practice we confine attention to two limiting cases, namely definitely distinguished, and entirely undistinguished particles. If the particles are definitely distinguished, $\chi = 0$ or π, and we can arrange that χ shall always be 0. If the particles are indistinguishable observationally, χ is an unobservable. In other words the rotation $q = e^{\frac{1}{2}Q\chi}$ is a relativity transforma-

tion to a different but equivalent frame of reference. This is an altogether new type of relativity transformation, which appears for the first time in double systems. But analytically it is of the same form as the well-known relativity rotations, and it has the same consequences. When χ is unobservable, the hamiltonian cannot contain χ explicitly, but it will contain the momentum (if any) conjugate to χ. That is to say, χ is an ignorable coordinate.

There is nothing mystical about the effects of indistinguishability. We do not suppose that an electron knows that it will not be distinguished from other electrons, and on that account conducts itself differently. We can imagine a being more gifted than ourselves who identifies each individual electron. He applies the ordinary equations of distinguishable particles to them, and his results are right; but his solutions do not interest us, because we can never obtain the observational data which he uses, and have no opportunity to apply or test his deductions. He observes, let us say, a red particle and a blue particle; he finds that the red particle is at x_1 at time t_1, and at x_2 at time t_2, and deduces that it had a velocity $(x_2-x_1)/(t_2-t_1)$ and a momentum $m(x_2-x_1)/(t_2-t_1)$. We observe a particle at x_1 at time t_1, and a particle at x_2 at time t_2, but we do not know whether it is the same particle. Velocities of particles (and the corresponding kinematical momenta) are not observational data for us; and a system of dynamics which manipulates such data is useless for our purposes.

The dynamical equations therefore depend, not on whether the particles are intrinsically distinguishable or indistinguishable, but on whether and to what extent they are in fact distinguished.

We must distinguish the general wave function $\Psi'(x,x',\chi)$ which may be any function of its arguments from the particular wave function $\Psi(x,x',\chi)$ defined in (15·22). The latter represents a uniform probability distribution in χ. If the states corresponding to different values of χ have different probabilities p_χ, their combined wave function is $p_\chi^{\frac{1}{2}}\Psi(x,x',\chi)$; or more generally we can form a combination with different probabilities of different states $\Psi(x,x')$ for each permutation angle.

Returning to the problem of the red and blue particles, the operators $-i\partial/\partial x$, $-i\partial/\partial x'$ do not give the momenta of the red and blue particles—which are the momenta referred to in the dynamical equations for distinguished particles. We can, if we like, regard $-i\partial/\partial x$ as the momentum of a composite particle, which has certain probabilities of being the red or blue particle. We have not the data required for applying ordinary dynamics, which does not profess to treat particles composite in this sense. But we can construct a formally similar dynamics by treating χ as an additional coordinate; so that the whole momentum of the system includes, besides the momenta of two composite particles, a momentum of interchange.

We have then to consider distributions of probability over the domain of x, x', χ, and more especially to determine distributions which represent steady states. When account is taken of the spin components of the stream vector, the distribution is represented in phase space; and the ordinary 136-dimensional phase space of a double system is extended to 137 dimensions by the addition of a dimension representing change of χ. We shall see later that the 137-dimensional phase space is a development of the "augmented phase space" used in § 12·3; so that from this point of view the permutation coordinate appears as compensation for the dropping of the relative time coordinate $t_2 - t_1$ when two systems are combined into one. If we have observational information—e.g. an observation of colour—which at any time distinguishes the particles with fairly strong probability, this will be represented as a wave packet in χ; just as observational information as to their position is represented by a wave packet in x, x'. The wave packet will gradually disperse—not because the particle is liable to change colour, but because we cannot trace definitely which of the particles whose colour we observed is the one which is now at x.

From the general problem we pass to the two special cases. Firstly, when the particles are entirely undistinguished, the equations are simplified by the fact that χ is an ignorable coordinate and can be eliminated, leaving only a term in the hamiltonian which represents its conjugate momentum. Secondly, when the particles are completely distinguished, the equations are again simplified because χ is constrained to be 0. This is more than a limiting case; it is a change in our point of view. We could not in an ordinary way prevent the wave packet concentrated at $\chi = 0$ from dispersing. The constraint involves a change in the basis of statistics. An astronomer may inadvertently interchange the two components of a double star which he is observing; but he treats this as a *mistake*. He does not expect the laws of celestial mechanics to predict his "observational result". The constraint $\chi = 0$ is imposed by stigmatising any other value of χ as a blunder. We have seen that the dynamics of distinguishable particles cannot from its very nature be applied to indistinguishable particles; but why should not the dynamics of indistinguishable particles apply (so far as it goes) to distinguishable particles? The answer is that theoretical physics is intended to agree with the experience of an observer who does not make mistakes, and will take varying forms according to the definition of what constitutes a mistake.

Analytically, the probability distribution of distinguishable particles is limited to the 136-dimensional section $\chi = 0$ of the 137-dimensional phase space. A change of the basis of statistics is made by limiting the initial probability distribution to 136 instead of 137 dimensions. Neglect of this distinction between the dynamics of distinguishable and indistinguishable

particles has caused an error of a factor $\frac{136}{137}$ in many of the currently accepted formulae of quantum theory.

From the present point of view *protons are indistinguishable from electrons*. It might be thought that the greater mass of the proton would be a distinction impossible to overlook. But the mass is determined by the operator $-i\partial/\partial t$. We cannot differentiate the probability distribution of a particular particle until we have settled how to identify that particle at different times. Thus mass can never be used as a criterion for distinguishing particles; it presupposes that they have already been distinguished.

The Principle of the Blank Sheet requires that at the start we should recognise no intrinsic distribution between the particles which we contemplate, in order that we may trace to their very source the origin of those distinctions which we recognise in practical observation. The fundamental dynamics is the dynamics of indistinguishable particles; the dynamics of distinguishable particles is a practical adaptation to be used when we do not wish to analyse the phenomena so deeply.

15·3. The Fermi-Dirac Law.

In current theory it is usual to take the wave function Ψ of two particles to be single-valued. Then, if Q is the interchange operator, $Q^2=1$. In place of (15·22) we take an extended wave function

$$\Psi(x, x', \chi) = e^{\frac{1}{2}i\chi(Q-1)}\Psi(x, x'). \tag{15·31}$$

Then $$e^{\frac{1}{2}i\pi Q} = \cos\tfrac{1}{2}\pi + iQ\sin\tfrac{1}{2}\pi = iQ$$

and $\Psi(x, x', \chi+\pi) = \Psi(x', x, \chi)$ as before.

Consider the tensor transformation

$$q = e^{\frac{1}{2}i\chi(-P-1)}, \tag{15·32}$$

where P is the interchange operator of the frames E_μ, F_μ of the two particles. Let their coordinates in three dimensions be x_μ, x_μ', the time t being common to these and other particles. Their combined position is specified by a position vector

$$X = \sum_1^3 (E_{\mu 4}x_\mu + F_{\mu 4}x_\mu') \tag{15·33}$$

in the double frame. We use the strain vector form of X, since the time is to be treated as invariant. A wave function of the two particles will be denoted indifferently by $\Psi(x_\mu, x_\mu')$ or $\Psi(X)$.

When the transformation (15·32) is applied, we have to consider, as in § 8·3, not only the direct change of Ψ but the change of its argument X. Usually the transformation will introduce new matrices into X, so that $\Psi(X)$ ceases to represent a distribution in the original 3-space. But when

$\chi=\pi$, we have $q=-P$; and the transformation of the strain vector X is (since P is space-like)

$$X'=\Sigma P(E_{\mu 4}x_\mu+F_{\mu 4}x_\mu')P$$
$$=\Sigma(F_{\mu 4}x_\mu+E_{\mu 4}x_\mu')$$

by (10·33). Thus we return to the original 3-space, but x_μ and x_μ' are interchanged. The transformed value of Ψ is

$$\Psi'=-P\Psi=\bar{\Psi}$$

by (10·37). Hence the combined result is

$$\Psi_{\alpha\beta}'(x_\mu, x_\mu')=\Psi_{\beta\alpha}(x_\mu', x_\mu), \tag{15·34}$$

Ψ' referring to the transformed distribution and Ψ to the original distribution. Using the permutation coordinate, we can also write (15·34) as

$$\Psi'(x,x',0)=\Psi(x,x',\pi)=e^{\frac{1}{2}i\pi(Q-1)}\Psi(x,x',0) \tag{15·35}$$

by (15·31).

Thus, in a sense, $Q-1=-P-1$; so that we have found a possible form for the hitherto unidentified symbol Q. The actual relation is

$$e^{\frac{1}{2}i\chi(Q-1)}=[e^{\frac{1}{2}i\chi(-P-1)}], \tag{15·36}$$

where the right-hand side is construed not as a simple multiplier but as a tensor transformation applied to the function which follows. (It is therefore not permissible to cancel out $e^{-\frac{1}{2}i\chi}$ on each side.)

Let us now treat χ as a dynamical coordinate, so that the transformation q gives an extended wave function representing a probability distribution over the seven-dimensional domain (x, x', χ). The six-dimensional sections $\chi=0$, $\chi=\pi$ coincide with ordinary space (repeated for the two particles). If the wave function is single valued in ordinary space, $\Psi(x,x',0)$ and $\Psi(x,x',\pi)$ must agree; so that by (15·35), $\Psi'=\Psi$. Hence, by (15·34),

$$\Psi_{\alpha\beta}(x,x')=\Psi_{\beta\alpha}(x',x).$$

Thus the wave function is symmetrical for interchange of the particles.

An antisymmetrical wave function is obtained by the transformation

$$q=e^{\frac{1}{2}i\chi(-P+1)} \tag{15·37}$$

which leads similarly to

$$\Psi_{\alpha\beta}(x,x')=-\Psi_{\beta\alpha}(x',x).$$

In this case (15·35) is replaced by $\Psi'(x,x',0)=-\Psi(x,x',\pi)$; so that the transformation (15·37) is not a simple rotation in χ. Relativistic rotation, or parallel displacement, in χ generates a special wave function, which we have called the extended wave function; this is symmetrical. But, as we have pointed out (p. 285), the general wave function may be any function of the coordinates x, x', χ, subject to the usual conditions of single-valuedness; and the antisymmetrical wave function is one of the general functions. Since it is not generated by parallel displacement in χ, its covariant de-

rivative with respect to χ does not vanish. Denoting the antisymmetrical wave function generated by (15·37) by Ψ^a, and distinguishing transformation operators from multiplying operators by square brackets, we have

$$\begin{aligned}\Psi^a(x, x', \chi) &= [e^{\frac{1}{2}i\chi(-P+1)}]\,\Psi(x, x') \\ &= [e^{i\chi}]\,[e^{\frac{1}{2}i\chi(-P-1)}]\,\Psi(x, x') \\ &= [e^{i\chi}]\,e^{\frac{1}{2}i\chi(Q-1)}\,\Psi(x, x') \\ &= [e^{i\chi}]\,\Psi(x, x', \chi). \qquad (15\cdot38)\dagger\end{aligned}$$

The covariant derivative of Ψ^a with respect to χ is contained in the factor $[e^{i\chi}]$, since the covariant derivative of $\Psi(x, x', \chi)$, which denotes as before the extended wave function formed by parallel displacement of $\Psi(x, x')$, is zero.

The following remarks will perhaps make the significance of the mathematical procedure clearer. Consider a series of six-dimensional states $\Psi(x, x', \alpha)$, distinguished from one another by a parameter α. In the special case in which α is unobservable this mode of division into states becomes degenerate, and the series can only be contemplated as a whole, i.e. as a seven-dimensional state. If α is promoted to the rank of coordinate, $\Psi(x, x', \alpha)$ is one such seven-dimensional state; but we can form other states $p(\alpha)\,\mathrm{X}(x, x', \alpha)$ by combining six-dimensional distributions with different (algebraic) probability coefficients $p(\alpha)$. In the seven-dimensional state the probability fluid (§ 8·1) is no longer restricted to flow in the planes $\alpha = \text{const.}$, and there will in general be a probability flux along the coordinate α, determined by the covariant derivative operator $-i\delta/\delta\alpha$ and therefore depending on $p(\alpha)$. This flux was not allowed for in the original equation of continuity of the probability fluid in six dimensions; thus X will not satisfy the same differential wave equation as Ψ. For given α, $\Psi(x, x', \alpha)$ is a self-contained six-dimensional state, but $\mathrm{X}(x, x', \alpha)$ is not. In our application α is the permutation coordinate χ; and the primary object of the investigation is to find the term in the wave equation for X, arising from flux in the χ direction and depending on $p(\chi)$, which does not appear in the wave equation for Ψ. To obtain its value we must determine $p(\chi)$. For an arbitrary angular coordinate α, the only limitation on $p(\alpha)$ would be that it must be periodic in α with period 2π; but exceptionally $p(\chi)$ has also to fulfil conditions at the *half-period*. We can interchange the particles, either by changing χ to $\chi+\pi$ in the wave function $p(\chi)\,\mathrm{X}(x, x', \chi)$ of the seven-dimensional state, or by interchanging x and x' in a six-dimensional section, and the two results must agree. Our analysis of this condition

† The analytical part of the investigation is interrupted at this point. There is a direct continuation of it, starting from (15·38), in § 15·7.

can be put in the following form. We first ensure by our definition of $X(x, x', \chi)$ that its covariant derivative with respect to χ vanishes; that is to say, a change of the argument χ represents only the effect of the relativity rotation provided by the indistinguishability of the particles. The test at the half-period then shows that, if $X(x, x', 0)$ is a symmetrical function of x, x', no additional factor $p(\chi)$ is required; but if it is antisymmetrical, we must insert $p(\chi) = [e^{i\chi}]$. Although $p(\chi)$ is here expressed as an operator, we shall find later that it reduces to an eigenvalue $e^{i\chi/137}$; so that the result is not inconsistent with our original limitation of $p(\alpha)$ to algebraic values.

Our inability to distinguish observationally the separate six-dimensional states permits the probability to flow freely between them—a flux which, if they could be distinguished, would be represented in a different manner, viz. by transitions. We do not here attempt to predict to what extent advantage will be taken of this freedom; but we discover that the existence of a cyclic flux in the new direction will betray itself by its effect on the wave function for any of the six-dimensional sections. In particular, we have calculated the cyclic flux which is indicated when the flow in a six-dimensional section is that represented by an antisymmetrical wave function.

It should be added that we have here adopted the solutions which give the smallest covariant derivative. There exist also solutions $p(\chi) = [e^{2ni\chi}]$ for symmetric wave functions, and $p(\chi) = [e^{(2n+1)i\chi}]$ for antisymmetric wave functions. The ultimate effect of taking $n = 0$ is that we shall determine the *minimum* charge of a particle.

According to the principle of Fermi and Dirac, the wave function of two elementary particles is antisymmetrical. Their theoretical treatment only went so far as to show that it must be either symmetrical or antisymmetrical; the choice of an antisymmetrical function depends on empirical considerations. We cannot at the moment go further than they did. Accepting their conclusion, the wave function for a pair of elementary particles must be taken to be Ψ^a.

We must now consider the "relativity of identity". We have stressed the fact that a simple wave function $\psi(x)$ yields nothing observable, since it contains no reference to any comparison object for measuring x. This objection is removed in the double wave function $\Psi(x, x')$, since one particle can serve as comparison object for the other. But the objection reappears in $\Psi(x, x', \chi)$, since there is no comparison object for χ. Just as we require two particles to provide observable differences $x_\mu - x_\mu'$, we require four particles—a quadruple wave function—to provide observable differences $\chi - \chi'$. An absolute permutation coordinate referred to an abstract frame has no observational meaning, *even when the particles are distinguishable.*

We may *observe* certain distinctions, but the *interpretation* which we place on these distinctions is not a matter of observation. We have temporarily shelved this difficulty by identifying the particles by a characteristic (colour) recognised in some supernatural manner; but actually colour might be changed by Doppler effects, and it is no more an absolute criterion of identity than position.† All we can do is to lay down a scheme of equations to be conventionally adopted as a criterion that χ is constant, and then measure changes of χ relative to this standard.

Standing waves and progressive waves afford an example of the relativity of identity. We have seen that the particle represented by a standing wave is of composite identity (§ 13·1). That depends on the usual view that a progressive wave represents a particle with single identity throughout all time. But it is equally possible to define the particles represented by standing waves as the real individual particles; then the progressive waves represent composite particles.

Accordingly a system of two particles must be referred to a comparison fluid described by double wave functions and involving a comparable permutation coordinate χ_0. The angles χ and χ_0 being referred to an abstract "frame of identity" are unobservable; but the probability distribution of the relative angles $\chi - \chi_0$, contained in the quadruple wave function of the double object system and double comparison fluid, may have observable characteristics. The partition of the relative permutation angle into "absolute" permutation angles referred to a frame is an analogous problem to that which led to the theory of the Riemann-Christoffel tensor.

From this point our investigation bifurcates into two problems. Firstly (§ 15·4–15·6), we calculate the interaction of two particles forming a hydrogen atom. Secondly (§ 15·7), we calculate the interaction between pairs of particles in any assemblage.‡ There is a considerable difference between the methods used; and the fact that both lead to the same value of the Coulomb energy is a useful check on the details of the calculation.

15·4. Interaction in the Hydrogen Atom.

We shall now determine the interaction between two undistinguished elementary particles which form a steady system. We know that, for a steady state to be possible, the particles must be of opposite sign. The system is therefore identified with a hydrogen atom.

† Consider a double star with equal components, having a very large orbital velocity in the line of sight. We can distinguish the two components by observing that one is red and the other blue. Half a period later we can again distinguish the red star and the blue star. But the stars thus identified do not obey the accepted laws of celestial mechanics. The accepted laws are obtained by identifying the blue star at time t with the red star half a period earlier.

‡ Atomic nuclei are not considered.

As in § 12·6, we re-resolve the hydrogen atom into an external and an internal particle of masses

$$M = \tfrac{136}{10} m_0, \quad \mu = \tfrac{1}{136} m_0. \tag{15·41}$$

The time direction is as usual taken to agree with the momentum vector of the external particle. Interchange of the proton and electron does not affect the external particle. We therefore confine attention to the internal particle of mass μ. Its coordinates ξ_1, ξ_2, ξ_3 are the relative coordinates of the proton and electron, and are reversed in sign by interchanging them. The change is most simply described by using angular coordinates, so that the displacements and momenta are $r d\theta_\mu$, $-i\partial/r\partial\theta_\mu$. Radial displacement and momentum are treated in the same way by letting $\theta_r = \log r$. The interchange is then equivalent to reversing the sign of r, leaving the angular variables unchanged.

It may be well to repeat that the interchange is purely subjective—as when an astronomer inadvertently interchanges the two components of a double star, and so publishes a position angle wrong by 180°. By such a mistake relative coordinates are given the wrong sign; only in treating indistinguishable particles we do not count it as a mistake, since there is no criterion for deciding which is the right sign.

Continuous interchange is represented by taking $r = r_0 e^{i\chi}$; or if necessary we may take $r = r_0 e^{Q\chi}$, where Q is any symbolic square root of -1.

The states of the internal particle will consist of probability distributions over the space coordinates ξ_1, ξ_2, ξ_3 and the permutation coordinate χ. There is no time coordinate in an internal state.

It would not be illegitimate to treat states in which the distribution is over ξ_1, ξ_2, ξ_3, only, with χ a constant for the state. It is purely a question of practical application. As explained in § 15·2 the resulting system of dynamics would be true, but useless to an observer who (owing to inability to distinguish the particles) could not obtain the data required for its application. Every angular coordinate θ has to be treated as we here treat χ; if the conditions are such that it becomes unobservable, a degeneracy is introduced, and the states which would have been distinguished by different values of θ are run together into a single state. The ultimate reason for this is that in physical applications we have to take account of transitions (due to external perturbations) between the states given by different values of θ; as θ approaches degeneracy these transitions become more frequent and the calculation becomes unmanageable; for completely degenerate θ an infinitesimal perturbation is sufficient to cause transitions, and the states can only be treated as a combined whole. Conversely, the transition causes only an infinitesimal perturbation of external systems and is therefore unobservable; if we cannot detect a transition, we cannot distinguish the states.

We have to picture, or represent symbolically, a rotation which changes r continuously into $-r$. For a picture we require an extra-spatial dimension σ through which r can turn. For symbolic treatment a matrix or symbol E_σ must be associated with σ in the same way that E_1, E_2, E_3 are associated with ξ_1, ξ_2, ξ_3. Owing to the absence of a time coordinate, a symbol E_4 is standing idle; this is available to represent a dimension perpendicular to E_1, E_2, E_3, and we therefore set $E_\sigma = E_4$. The question whether E_σ is *really* the missing E_4 does not arise; it is sufficient that its commutation relations with all other symbols in the formulae are identical with those of E_4.† If E_r is the matrix associated with the direction of r when $\chi = 0$, a point in the four dimensions will be represented vectorially by $E_r r + E_4 \sigma$, and the χ rotation will be $q = e^{\frac{1}{2}E_4 E_r \chi}$.

The spread of the probability distribution over the additional coordinate σ introduces a corresponding term $E_4 \partial/\partial\sigma$ in the differential wave equation. Or, using the standard form (8·631) for a strain vector, we have

$$\left(E_{14}\frac{\partial}{\partial\xi_1} + E_{24}\frac{\partial}{\partial\xi_2} + E_{34}\frac{\partial}{\partial\xi_3} + \frac{\partial}{\partial\sigma} + iE_{45}\mu\right)\psi = 0. \qquad (15{\cdot}42)$$

The extra permutation dimension thus plays the same part in an internal state as the time dimension plays in an external state.

Except that we see that the energy of an internal state arises from interchange, being conjugate to the linear interchange coordinate σ, our only result thus far is to obtain the form of wave equation, already familiar for external particles. The equation expresses the conservation of probability; but the coordinates, etc., contained in it are unobservable, being referred to the frame and not to the comparison fluid. We have next to derive from it an equation suitable for practical use.

15·5. The Fine Structure Constant.

It is necessary to consider the internal particle in conjunction with its comparison fluid. Since $m_0 = 136\mu$, the observable angular displacements θ' are analysed into a displacement $\theta = \frac{136}{137}\theta'$ of the particle and a recoil $\theta_0 = \frac{1}{137}\theta'$ of the partial comparison fluid. The question now arises whether the same partition applies to displacements of the permutation coordinate χ. We see immediately that the partition of χ is governed by quite different considerations from the partition of θ which is based on the theory of the Riemann-Christoffel tensor.

In the first place there is no need to admit *any* recoil of χ. The comparison fluid is an idealised substitute for actual reference objects; and we are free to choose either a "distinguishable" or an "indistinguishable" comparison

† The use of rectangular relative coordinates implies that the region considered is small enough to be treated as flat, so that E_5 also is idle. But this absence of E_5 is a casual feature of a special problem; it is not comparable with the enforced absence of E_4, arising from the definition of an internal state as a simultaneous state.

fluid, i.e. fluids whose constituent particles are respectively distinguishable or indistinguishable. Inasmuch as the fluid replaces distinguishable macroscopic bodies, it is a less violent distortion of the natural conditions to use a distinguishable comparison fluid. Then χ_0 is constrained to be zero, and there is no recoil.

There is no breach with our previous conventions and definitions in admitting this exception, though it will complicate subsequent applications. Initially we were free to specify the recoil in each coordinate arbitrarily; but later we introduced the condition that the (negative) recoil in the time direction is such that the object particle and the comparison fluid move forward together in time. Relativity conditions then determine the recoil in any other direction which can be connected with the time direction by a relativity rotation. But the direction χ is defined by a tensor of different rank, and is not connected with the other directions by a relativity rotation.† It would be more of a breach with our previous conventions to admit an indistinguishable comparison fluid; for in all previous references to the comparison fluid we have treated it as composed of distinguishable particles. In particular its connection with the metrical tensor $g_{\mu\nu}$ has been fixed on that basis. To substitute an indistinguishable comparison fluid at this stage would derange the metric, and upset our standard equations.

Nevertheless it is instructive to consider what would be the result of employing an indistinguishable comparison fluid. In actual phenomena, shown by a comparison of the microscopic object particles with macroscopic reference objects, the comparison fluid is eliminated, and the final results are unaffected. The use of an indistinguishable comparison fluid as intermediary will simplify the microscopic part of the theory at the expense of complicating the macroscopic part of the theory of these phenomena. The macroscopic theory is affected because the energy tensor of the comparison fluid determines the metric to which the actual measures of length are supposed to be referred. We have seen that continuous interchange is represented by the transformation $r \to re^{q\chi}$. This transformation may be attributed to an absolute change of r, or to an absolute change of the standard of length. (Since the other variables are angular, r is the only quantity affected by a change of the standard of length.) The standard of length for the object system is contained in the comparison fluid—the idealised substitute for metre rods, etc. Hence the partition of χ' into χ and χ_0 corresponds to a partition of the apparent change of r into absolute changes of r and of the standard gauge. That is to say, the "recoil" of the change of r is a gauge transformation.

† We now propose to represent it as relativistically connected in a four-dimensional picture, assigning to it the matrix $E_4 E_r$; but the condition of admissibility of this representation is the point we are now considering.

For a recoil $\chi_0 = \frac{1}{137}\chi'$, the gauge is modified by the factor $e^{Q\chi'/137}$ and becomes complex (if $Q = i$) or matricised. If we are to pursue this method we must abandon Riemannian space and adopt Weyl's geometry which admits complex gauge transformations, or the author's extension of it which admits more general transformations of gauge equivalent to matrix transformations. In wave mechanics we prefer to keep to Riemannian space; then χ can have no recoil; and the equations of the microscopic theory are complicated by the fact that χ behaves differently from the other angular coordinates in this respect. Before studying the phenomena which result, we notice two points of interest:

(*a*) The equivalent complex gauge transformation corresponds to an electromagnetic field (§ 8·8). Thus the phenomena will be of electromagnetic character.

(*b*) The gauge would have been changed by a factor $e^{i\chi'/137}$ (or by a matrix having this eigenvalue); thus a coefficient 137 is introduced. This coefficient is known empirically as the *fine-structure constant*.

We have seen that formally σ takes the place of an imaginary relative time ξ_4. We shall try to elucidate this connection. Let (x_1, x_2, x_3, t), (x_1', x_2', x_3', t') be the coordinates of the proton and electron, and let $(\xi_1, \xi_2, \xi_3, \tau)$ be the differences $x_1' - x_1$, etc. By the definition of a combined system τ is constrained to be zero. We may, however, regard a displacement $d\tau$ as a transformation to a frame of reference with a different reckoning of simultaneity. Consider the transformation

$$r = r_0 \cosh u, \quad \tau = r_0 \sinh u.$$

The *interval* from the origin to the point considered is $(r^2 - \tau^2)^{\frac{1}{2}} = r_0$; and we can show easily that the intervals between all other pairs of points are likewise independent of u. Thus a physical system occupying the domain (ξ_1, ξ_2, ξ_3) is intrinsically unaltered by the transformation. Apparently all distances in it are expanded in the ratio $\cosh u$; but this is accounted for by the fact that a new reckoning of simultaneity has been introduced, which antedates each particle by a time $\tau = r \tanh u$, proportional to its distance from the origin.

The corresponding transformation for imaginary relative time σ is

$$r = r_0 e^{i\chi}, \quad \sigma = r_0 e^{i(\chi - \frac{1}{2}\pi)} = -ir.$$

For small displacements from the zero state ($u = 0$, $\chi = 0$) we have

$$d\tau = r_0 du, \quad d\sigma = r_0 d\chi.$$

Thus far our formulae refer to changes of the system of reference—real or imaginary changes of reckoning of simultaneity. But we can employ the transformation in the usual way to define a series of states of a system in the same frame of reference. The states will represent the same system (ξ_1, ξ_2, ξ_3)

expanded in the ratios $\cosh u$ and $e^{i\chi}$ respectively, the latter expansion being interpreted as in § 15·1, as a change of relative probability of the direct and interchanged states. The u series of states is unclosed (hyperbolic transformation) and a distribution of probability over it could not be represented in phase space. That is to say, in analysing the total distribution of probability over the domain (ξ_1, ξ_2, ξ_3), into the sum of a number of elementary distributions, we cannot accept a dissection into a series of states continuously varying in scale. The corresponding expansion of the wave function in a series of elementary wave functions would not be convergent. For that reason, when once we have fixed our frame of reference, the u transformation is excluded, and τ is identically zero. But in saying that an internal state is formed by the simultaneous configuration of two particles, we refer to simultaneity in real time; they may be non-simultaneous in imaginary time—if we can find a meaning for such a phrase. Thus the χ series of states is admissible (as a mode of dissection of the total probability distribution); and since it is a closed series (circular transformation) a distribution of probability over it can be represented in phase space.

This clears up a point which personally I have found most difficult. It had seemed to me that for a real displacement dr, just as for a displacement $d\chi$, there would be no recoil of the comparison fluid.† An increase of r to $r+dr$ might be attributed to an absolute change of r or to an absolute change of the standard of length, the latter constituting the "recoil" of the change of r. But since Riemannian geometry excludes changes of gauge, we can admit no such recoil. The argument is the same as that which led us to exclude recoil of χ which would represent imaginary change of gauge. In short, Riemannian geometry requires a comparison fluid rotatable in every direction, but not expansible or subject to interchange of its particles.

To see the fallacy of this argument, we must recall that the same displacement $d\xi_\mu$ may be produced by a variety of rotations in different planes, corresponding to different transformations of the wave vector. Therefore when we treat a displacement dr or a momentum $-i\partial/\partial r$, we must not jump to the conclusion that it corresponds to a u transformation. By reference to the mode of formation of the wave equation, we see that the r-momentum which appears in it is the E_{r5} component of a space vector, and corresponds to rotation about the centre of curvature of space-time; this, of course, produces the usual recoil of the comparison fluid as provided for in the R.C. tensor.

The nature of a displacement $d\xi_\mu$ at any point—whether it is a rotation about the centre of space-time, or about an origin in the domain (ξ_1, ξ_2, ξ_3), or an expansion of scale—cannot be immediately seen by inspection. It

† If the argument were correct, the Coulomb energy term found in § 15·6 would be duplicated.

depends on the system of analysis into elementary states furnished by the four commuting operators W, U_1, U_2, U_3. But without entering into the details of this analysis, we can show that there will be only one coordinate χ which produces no recoil. The argument above was that whatever applies to the imaginary transformation of r should also apply to the real transformation of r; but this is forestalled by the earlier result that of two antithetic transformations one will be circular and the other hyperbolic, and only the circular transformation is admitted in our system of analysis into elementary states. Thus we can only be concerned with one of the two transformations.

We may recall that an $E_{\mu 5}$ transformation displaces positive and negative charges in opposite directions in neutral space-time, and is therefore associated with polarisation rather than simple translation (§ 6·3). It is this aspect of the r displacement—as a change of separation of positive and negative charges—which is handled in the internal wave equation.

15·6. The Coulomb Energy.

In the wave equation (15·42) the internal particle is represented by a simple wave vector ψ. The justification for separating ψ from the double wave vector of the particle and comparison fluid, and treating it as an independent distribution, depends on the theory of Chapter XII; but no account was then taken of the permutation coordinate. To rectify this we must go back to the quadruple wave function of the proton and electron and the corresponding unspecified particles which form their comparison fluid. This is resolved into two interchangeable double wave functions; or into a double wave function of an external particle (without permutation coordinate) and comparison fluid, and a double wave function of an internal particle (with permutation coordinate) and comparison fluid. Considering the latter, the permutation coordinate raises the number of dimensions of the phase space to 137. If it behaves symmetrically with the other dimensions (which implies that there is the same recoil in χ as in the other coordinates) the same theory of resolution into simple wave vectors applies, except that the ratio of the mass μ' of the internal particle to the mass m_0 of the comparison fluid is now given by

$$137\mu' = m_0 \qquad (15{\cdot}61)$$

instead of $136\mu = m_0$.

Hence, if the comparison fluid recoils in χ, the wave vector ψ_r satisfies the wave equation of the form (15·42)

$$\left\{E_{14}\frac{\partial}{\partial\xi_1} + E_{24}\frac{\partial}{\partial\xi_2} + E_{34}\frac{\partial}{\partial\xi_3} + \frac{\partial}{\partial\sigma} + iE_{45}\mu'\right\}\psi_r = 0. \qquad (15{\cdot}62)$$

But our standard comparison fluid does not recoil in χ, and we wish to find the wave equation of the corresponding wave vector ψ.

In (15·62) a displacement $d\sigma = ir\,d\chi$ is accompanied by a recoil $d\chi_0 = \frac{1}{137}d\chi$ of the comparison fluid, in accordance with the relation $m_0 d\chi_0 = \mu' d\chi$. To eliminate this recoil we must transform the frame of reference so that the permutation coordinate of the comparison fluid in the new frame remains constant. The required rotation of the frame is in the forward direction of χ_0, since the "recoil" is really an advance (owing to the time-like character of the matrix associated with it). The transformation due to the rotation of the frame is therefore

$$\begin{aligned}\psi &= e^{-\frac{1}{2}i\chi_0}\psi_r \\ &= e^{-\frac{1}{2}\sigma/137r}\psi_r, \text{ for small values of } \sigma. \end{aligned} \tag{15·63}$$

The transformation matrix is properly the matrix F_{r4} associated with the χ_0 rotation of the comparison fluid in its own frame F_μ; but, since F_{r4} commutes with all the symbols in (15·62), it is for our purposes an algebraic square root of -1. Substituting in (15·62), we have

$$\left\{E_{14}\frac{\partial}{\partial\xi_1} + E_{24}\frac{\partial}{\partial\xi_2} + E_{34}\frac{\partial}{\partial\xi_3} + \frac{\partial}{\partial\sigma} + iE_{45}\mu'\right\}e^{\frac{1}{2}\sigma/137r}\psi = 0,$$

so that (putting $\sigma = 0$ after the differentiation)

$$\left\{E_{14}\frac{\partial}{\partial\xi_1} + E_{24}\frac{\partial}{\partial\xi_2} + E_{34}\frac{\partial}{\partial\xi_3} + \frac{\partial}{\partial\sigma} + \frac{1}{2}\cdot\frac{1}{137r} + iE_{45}\mu'\right\}\psi = 0, \tag{15·64}$$

which is the required wave equation.

Comparing with (9·64), we have

$$\frac{\pi e^2}{hc} = \frac{1}{2}\cdot\frac{1}{137}.$$

Hence the value of the fine structure constant is determined as

$$hc/2\pi e^2 = 137. \tag{15·65}$$

15·7. Interaction in Systems of Particles.

We turn now to the general problem of interaction in an assemblage of elementary particles. The total interchange energy of a particle will arise from interchange with every other particle. The difference from the previous investigation is that, since the particle has to be paired with more than one other particle, the special coordinates x_μ, ξ_μ are inappropriate; and we have to obtain expressions in terms of the coordinates x_μ, x_μ' referred to an arbitrary origin.

Considering a wave function Ψ of two particles, we resume the investigation in § 15·3 at equation (15·38), where it was broken off. In that equation the operator $e^{\frac{1}{2}i\chi(Q-1)}$ or $[e^{\frac{1}{2}i\chi(-P-1)}]$ gives the change of Ψ due to parallel displacement in χ. (Displacement in χ is, for indistinguishable particles, a relativistic rotation of the frame of identification, and these operators give

the nominal change of Ψ due to its being referred to the rotated frame.) For antisymmetrical wave functions there is by (15·38) also a "real change" of Ψ represented by the operator $[e^{i\chi}]$. The covariant derivative $\delta/\delta\chi$ measures the real change; that is to say, $\delta/\delta\chi$ for Ψ is equal to $\partial/\partial\chi$ for $[e^{i\chi}]$. But $[e^{i\chi}]$ is not a simple multiplier; it turns $\Psi(X)$ into $e^{i\chi}\Psi(e^{2i\chi}X)$. We shall show that the change of argument eliminates $\frac{136}{137}$ths of the factor $e^{i\chi}$.

When $X \to e^{2i\chi}X$, the coordinates become $e^{2i\chi}x_\mu$, $e^{2i\chi}x_\mu'$. Since Ψ represents a probability distribution over real coordinates, its interpretation for a complex argument requires definition. We are interested only in the covariant derivative of $[e^{i\chi}]$; therefore any part of it which represents parallel displacement of the function $\Psi(x, x', \chi)$ on which it operates may be ignored. A parallel displacement is produced by applying a tensor transformation $q = e^{i\alpha}$ to all the symbols concerned; this yields

$$e^{i\alpha}\Psi(e^{2i\alpha}x, e^{2i\alpha}x', \chi'),$$

where χ' is the transformed value of χ, whose tensor character we leave for the moment undetermined. We therefore define Ψ for a complex argument of the form $e^{2i\alpha}X$ by

$$e^{i\alpha}\Psi(e^{2i\alpha}x, e^{2i\alpha}x', \chi') = \Psi(x, x', \chi), \qquad (15{\cdot}71)$$

since the two expressions are in any case equivalent for the purpose required. We want, however, to find $\Psi(e^{2i\alpha}x, e^{2i\alpha}x', \chi)$; for α will ultimately be put equal to χ, and we naturally do not apply the transformation to the argument which defines the transformation.†

The transformation of the double strain vector of the distribution can be treated as a uniform gauge transformation $R \to Re^{i\beta}$ of the radius of the 137-dimensional phase space. Denoting the strain vector $\Psi\Phi^*$ by S, the first effect of the transformation is to multiply S by $e^{137i\beta}$ on account of the change of measure of the volume element attached to it; and the second effect is to multiply it by $e^{-137i\beta}$ on account of the renormalisation necessary through the whole volume of phase space becoming $e^{137i\beta}\Omega$. These are equivalent to the two compensating changes in (15·71). It does not much matter which way round we identify them; but presumably the first is the result of the change of coordinates in the argument, and the second is the direct tensor transformation of Ψ and Φ by a factor $e^{i\alpha}$, which gives a factor $e^{2i\alpha}$ in S. Thus $2\alpha = -137\beta$.

If we exempt the coordinate χ from the transformation, so that the volume element contains $d\chi$ instead of $e^{i\beta}d\chi$, the first factor becomes $e^{136i\beta}$. The exemption is a simple matter, because the interchange matrix P commutes with all the space-like matrices in the CD-frame, i.e. the frame in

† That would lead to the same confusion as the statement "every number on this page should be divided by 2", which (being amended in accordance with its own instruction) implies that the numbers should be divided by 1, and therefore that they should be divided by 2, etc.

which the symmetrical and antisymmetrical wave functions are separated (§ 10·6). The second factor $e^{-137i\beta}$ is unaffected. The complete cancelling only occurs when the transformation (parallel displacement) is applied to *all* the quantities concerned; in withdrawing one of them from the transformation, we do not have to introduce any compensating change elsewhere. The resultant change of S is therefore $e^{136i\beta}.e^{-137i\beta}=e^{2i\alpha/137}$. The change of Ψ or Φ is $e^{i\alpha/137}$. Accordingly, putting $\alpha=\chi$,

$$[e^{i\chi}]=e^{i\chi/137}. \tag{15·72}$$

Hence the angular momentum is

$$-i\delta/\delta\chi=1/137. \tag{15·73}$$

To summarise the argument from the beginning: Indistinguishability of two particles causes a degeneracy, which is treated (like other cases of degeneracy) by introducing a dynamical coordinate χ. The angular momentum $-i\,\delta/\delta\chi$ is quantised by the condition that Ψ shall be single-valued. The change of Ψ with χ may be simple parallel displacement, in which case there is no angular momentum;† or there may be an additional real change. But unlike other cases of quantisation, the condition of single-valuedness is applied after a half-period when the original distribution has reappeared with the particles interchanged; and the Fermi-Dirac condition determines the real change to be that expressed by the transformation operator $[e^{i\chi}]$. If a transformation $[e^{i\alpha}]$ were applied to $\Psi(X,\chi)$, χ being transformed in accordance with its tensor character as well as Ψ and X, the result would be a parallel displacement which contributes nothing to the angular momentum. In the present case the transformation $[e^{i\chi}]$ is not applied to the coordinate χ whose value it fixes. Consequently, instead of the factor $e^{i\chi}$ being completely neutralised by the change of argument of Ψ, 1 out of the 137 coordinates of phase space is left untransformed, and fails to do its share in neutralising $e^{i\chi}$. Thus a factor $e^{i\chi/137}$ survives.

We have now to insert in the wave equation an extra term representing the linear momentum which corresponds to (15·73). We recall that the linear coordinate σ corresponding to the χ rotation is the imaginary relative time of the two particles. So that, if r is the distance between the two particles, $d\sigma=r\,d\chi$. Let iK_r be the matrix (in the double frame) associated with the direction of r.‡ It will be a γ matrix, which by (10·35) commutes with P, and is therefore unaltered by interchanging the particles and measuring r in the reverse direction. The matrix of the χ rotation, given by (15·32), is $-i(P+1)$. The matrix of the direction of σ is the product $K_r(P+1)$.

† Other than the $\frac{1}{2}$ quantum, which is added in our usual non-relativistic treatment, and dropped again when the angular momentum is converted into linear momentum.

‡ We take iK_r so that the direction may be associated with a symbol whose square is -1, as in a simple frame. It is understood that the strain vector representation is used throughout.

The new momentum to be inserted in the wave equation is therefore (cf. (8·37))

$$K_r(P+1)\left(-\frac{i\delta}{r\delta\chi}\right)=\frac{K_r(P+1)}{137r}. \qquad (15\cdot74)$$

This is divided into two parts. The term $K_r/137r$ embodies the reduction from the metric naturally associated with indistinguishable particles, whose initial probability distribution extends over 137 dimensions, to the standard metric associated with distinguishable reference objects. If there are two particles only, one of which is taken as the origin of polar coordinates, so that the hamiltonian contains a term $(iK_r)(-i\partial/\partial r)$, we can amalgamate $K_r(\partial/\partial r+1/137r)$ into $K_r\partial/\partial r$ by setting $\Psi=r^{-1/137r}\Psi'$, and taking Ψ' as the new wave function (cf. the treatment of (8·38)). But since we have already treated the case of two particles more directly, it is unnecessary to pursue this further. It is sufficient to notice that the term $K_r/137r$ is the equivalent in the general formula of the change of mass from μ to μ' in the internal wave equation of the hydrogen atom, which will be considered further in the next section. The remaining term represents the Coulomb energy:

$$\text{Coulomb energy}=\frac{K_rP}{137r}. \qquad (15\cdot75)$$

The main additional result is that the matrix associated with the Coulomb energy is K_rP. We naturally associate the Coulomb energy (or more strictly the Coulomb momentum) with the direction of r; we now see that it is necessary to multiply the matrix giving the direction of r by the symbol P in order to obtain the direction of the Coulomb momentum. We may therefore describe the Coulomb momentum as equal to $P/137r$ in the direction of r. In current perturbation theory it is taken to be $P_1/137r$, where P_1 is given by (10·385), the other factor of P being omitted. How far this omission is justified will depend on the nature of the application.

Light is thrown on the occurrence of P by the transformation investigated in § 10·8. On multiplying matrices in the EF-frame by P, we obtain matrices in the GH-frame. Thus the factor P disappears when we adopt the GH-frame, i.e. when we divide the system of two particles into external and internal wave functions. Conversely we might have anticipated that the purely algebraic Coulomb energy, found in the internal wave equation of the hydrogen atom, would acquire a matrix P when referred to the frame of the coordinates x_μ, x_μ'.

The previous investigation treated the interaction between particles of unlike sign; the present investigation applies most obviously to particles of like sign. We must suppose that E_μ and F_μ are both right-handed or both left-handed frames; otherwise there is no interchange operator P. But since no assumption has been made as to the nature of the stream vectors of the two particles, they may well have charges of opposite sign. It would, I think,

require deeper analysis to show that the sign of the energy depends on whether the charges are like or unlike. In this section we have been content to use the Fermi-Dirac principle (partly theoretical and partly empirical) to obtain the operator $[e^{i\chi}]$; the principle is equally consistent with $[e^{-i\chi}]$, so that the sign of the energy is left undetermined. It is possible, however, to go behind the Fermi-Dirac principle and, by investigating the quadruple wave function, trace the Coulomb energy to the absence of recoil of the χ coordinate in the comparison fluid, as we have done in the special case of a hydrogen atom. I think it would not be difficult to verify in this way the dependence of the energy on the sign of the charges.

15·8. The Factor $\frac{137}{136}$.

A new point brought out in this investigation is that the mass concerned in the internal wave equation is not μ but μ', where, by (15·61),

$$\mu' = \tfrac{136}{137}\mu. \qquad (15·81)$$

In considering the consequences of this, we must remember that μ does not represent an energy actually present. The mass of a hydrogen atom on the verge of ionisation is M; and although, for the purposes of analysis, an internal mass μ or μ' is added, it is subtracted again at the end of the investigation. When the atom is in a lower quantum state, μ is involved as a coefficient in the energy tensor $\mu^{-1}\delta^2/\delta\xi_\beta\delta\xi_\alpha$ of the internal wave function. When μ is reduced to μ' the energy tensor is increased, and the energy differences between the different states are increased. The actual energy differences are therefore $\frac{137}{136}$ times greater than those calculated in the usual way from Sommerfeld's formula (9·372), which ignores this factor.†

It will be seen that the decrease of μ has just the opposite effect to that which we should at first have expected. The explanation is that μ is not under any circumstances a rest mass, but occurs only as a *divisor* in the expression $\frac{1}{2}(\varpi_1^2+\varpi_2^2+\varpi_3^2)/\mu$ for the kinetic energy. In particular it is not permissible merely to substitute μ' for μ in Sommerfeld's formula.

The following alternative treatment leads to the same result. Consider a proton and electron confined within a rectangular boundary, so that there is no angular momentum, and the steady state can be analysed into standing waves in three perpendicular directions. We take the energy to be either positive or so slightly negative that the size of the atom is comparable with the dimensions of the enclosure. Equating the energies of the proton and electron to the energies of the external and internal particles, we have, by (10·97),

$$\left(m_p+\frac{p_p^2}{m_p}\right)+\left(m_e+\frac{p_e^2}{m_e}\right)=\left(M+\frac{P^2}{M}\right)+\left(\mu+\frac{\varpi^2}{\mu}\right)-\mu. \qquad (15·821)$$

The Coulomb energy is omitted.

† In (9·372) the symbol μ is used for a quantum number, and the internal mass $m_p m_e/(m_p+m_e)$ is denoted by m.

Subtract from each term the energy m_0 of a comparison particle, so as to leave only the energy due to specification. Then

$$\left(m_p+\frac{p_p^2}{m_p}-m_0\right)+\left(m_e+\frac{p_e^2}{m_e}-m_0\right)=\left(M+\frac{P^2}{M}-m_0\right)+\left(\mu+\frac{\varpi^2}{\mu}-m_0'\right), \tag{15·822}$$

where
$$m_0'=m_0+\mu=\tfrac{137}{136}m_0 \tag{15·83}$$
by (15·41).

Thus in order that the external and internal energies may be treated as additive in the same way as the energies of the proton and electron, the internal energy must be referred to a comparison mass m_0' which is $\frac{137}{136}$ times the comparison mass for external wave functions. The current treatment replaces $(\mu+\varpi^2/\mu-m_0')$ by a linear hamiltonian, taking it for granted that the theory developed for external wave functions applies. We have seen that it is necessary to decrease μ to μ', which is equivalent to increasing the comparison standard from m_0 to m_0', so far as the internal state is concerned. But when we introduce the condition that the energy of specification of the atom shall be the sum of the energies of specification of its internal and external states, (15·822) shows that the right course is, not to diminish μ to μ' in the current (Sommerfeld) formula, but to treat the formula as expressing the energy in a unit which is m_0'/m_0 times the unit used for the external energy.

Ideally we can measure experimentally the energy $\epsilon-\mu$ of a particular quantum state of the hydrogen atom, and hence determine the mass μ by Sommerfeld's formula. The external mass M can be found experimentally by some procedure equivalent to counting the number of atoms in a quantity great enough to be weighed macroscopically. From M and μ, m_p and m_e can be calculated. But the ordinary determination of the masses by this method, which omits the factor $\frac{136}{137}$, will be incorrect. The observed energies $\epsilon-\mu$, or the observed energy differences between different quantum levels, must be decreased in the ratio $\frac{136}{137}$ before being compared with Sommerfeld's formula.

Since M is very nearly the mass of the proton and μ is very nearly the mass of an electron, the practical effect of neglecting the factor is that the so-called observed mass of the electron is $\frac{137}{136}$ times too great, that of the proton being correct. It was first pointed out by W. N. Bond† that the observed values of the various constants would come into line with the author's theory, if it could be assumed that the observational determinations of e/m_e really determined $\frac{136}{137}e/m_e$. At that time I was aware that a factor $\frac{136}{137}$ would be involved, but had not been able to determine its precise incidence.

† *Nature*, **133**, 327 (1934).

More precisely the uncorrected observed determinations of m_p and m_e should be the roots of

$$10m^2 - 136mm_0 + \tfrac{137}{136}m_0^2 = 0. \qquad (15{\cdot}84)$$

Their ratio is 1834·1.

We cannot assume that the factor $\frac{136}{137}$ will occur in, or have the same incidence in, all methods of determining m_p/m_e or e/m_e. The theory of each method should be examined in detail in the light of what we have learned. In the next section, we shall express our result in a different form, which shows the range of its observational consequences more clearly.

15·9. Revision of the Constants e, m_p, m_e, h.

In saying that 1847·6 is the correct value of the mass-ratio, and that the "observational value" 1834·1 is in error through neglect of a factor in the reductions, we are employing the definition of mass or energy which has been regarded as fundamental in wave mechanics, namely as the value of the operator $(-ih/2\pi)\,\partial/\partial t$, h being a universal constant. Naturally we do not guarantee that 1847·6 will be the correct value if mass is defined in some other legitimate, and perhaps preferable, way. It appears that the determination of e/m_e by the deflection method leads to the mass-ratio 1834·1. It cannot well be supposed that the factor $\frac{136}{137}$ is concerned in this case. But the deflection method determines a mass which satisfies the classical definition. It is therefore not necessarily in conflict with our determination of the mass-ratio according to the quantum definition; though it has an important bearing on the larger question—whether for general purposes the quantum definition of mass is the best to adopt.

The discovery of the factor $\frac{136}{137}$ creates a new situation; and there can, I think, be no doubt that the most satisfactory way of restoring order is to admit two constants h, h' to be used in connection with internal and external wave functions respectively. The remaining constants e, m_e, m_p have unique values. This is a radical change. From our first introduction to quantum theory we have been taught to regard $E = h\nu$ as its most inviolable principle. But we did, in fact, tacitly abandon it, when we were forced to recognise that the momentum operator depends on the index of the wave function, and that the index of Dirac's Lorentz-invariant wave function is not the same as that of the commonly used wave function of the hydrogen atom. Moreover, if any definition has to be altered, h is the obvious victim; because, unlike e, m_p, m_e, it does not occur in classical theory.

We take $$h' = \tfrac{136}{137}h.$$

Then, if the double-valued internal wave function with operator $(-ih/\pi)\,\partial/\partial x$ is taken to be of index 1, the double-valued external wave function with operator $(-ih'/\pi)\,\partial/\partial x_\mu$ is of index $\frac{137}{136}$.

The two fine-structure constants are

$$\alpha = hc/2\pi e^2 = 137, \quad \alpha' = h'c/2\pi e^2 = 136.$$

The fine-structure constant is merely a name for the number of dimensions of the double phase space concerned.

As compared with our previous results, external masses will be reduced in the ratio $\frac{136}{137}$, and internal masses will be unchanged. Counting m_0 as an external mass, our previous results become

$$M = \tfrac{136}{10} m_0, \quad \mu = \tfrac{137}{136} \cdot \tfrac{1}{136} m_0.$$

Hence m_p, m_e satisfy (15·84), and the mass-ratio is 1834·1. Bond's correction has been eliminated by the change in the definition of mass. It was originally required to reduce internal energies to the unit which was used for external energies; but we have now redefined external energies in a way equivalent to changing the unit.

Let us consider the deflection method of determining mass. It would ideally be possible to determine the ratio m_p/m_e by comparing the deflections of a proton and electron projected with known energy in the same magnetic field. The motion will be in accordance with the wave equation

$$(\Sigma E_\mu p_\mu - m)\psi = 0, \quad \text{with} \quad p_\mu = (-ih'/2\pi)\,\partial/\partial x_\mu + \kappa_\mu,$$

the wave functions being external. As h' will not appear in the ratio of the two deflections, the fact that it differs from h will not be revealed. The ordinary calculation will therefore give the correct mass-ratio, which on the new system is 1834·1.

So far as radiation is concerned, the normal constant h is applicable; for, in absorption and emission, radiation is connected with the quantised, i.e. internal, states of material systems.

Thus in general the uncorrected observational determinations of the natural constants should be consistent with $\alpha = 137$, $m_p/m_e = 1834{\cdot}1$. Either they depend only on classical theory, and do not introduce h; or they depend on internal wave functions or radiation frequencies. To introduce h' we require an experiment determining the absolute wave length of the external wave functions of electrons or protons. Presumably the diffraction of electrons by matter involves h'; and the ordinary calculation will give the scale of the diffraction pattern $\frac{137}{136}$ times too large, unless the effect is concealed by a compensating factor. At present it is not possible to attain this accuracy. It is difficult to devise any other experiment in which the factor could manifest itself.

Summarising our conclusions—we find that the definition of energy in wave mechanics by the formula $E = h\nu$ is not in all cases consistent with the established meaning of the term "energy" in classical theory. In this section we restore the classical reckoning with e, m_p, m_e as fundamental constants. We calculate the ratio m_p/m_e to be 1834·1. The ratio e/m_p or e/m_e must be

found experimentally since it involves our arbitrary standards of mass and length. It can be found from the Faraday constant, by the oil-drop method, or by deflection methods, none of which involve h. We must further consider how results that have been expressed in terms of h are to be reduced to determinations of e, m_p, m_e. We first, by using $\alpha = hc/2\pi e^2$, express them in terms of α instead of h. Then α denotes the number of dimensions of the double phase space of the wave function concerned and its comparison fluid, and may be either 137 or 136. In the experiments which furnish the most accurate values of the constants the value is 137; but the theory of each experiment should be scrutinised with this point in view.

This summary covers most experimental determinations of the constants; but there is one important experiment which introduces new considerations. We shall treat it in the next section.

15·95. The Crystal Grating.

It is now considered by leading authorities that the various observational methods of determining e, m_p, m_e, h cannot be completely reconciled by current theory. The discrepancy, as usually stated, is that the value of e found by the crystal grating method is definitely greater than that found by other methods. To put the blame on e is merely conventional; the discrepancy is in the relations of the constants, and can equally well be attributed to m_p. The question naturally arises whether this discrepancy has anything to do with the factor $\frac{136}{137}$.

We must first exhibit the disagreement in a form free from irrelevancies. The crystal experiment consists in comparing the diffraction of the same X-rays by (*a*) a ruled diffraction grating, and (*b*) a natural diffraction grating formed by the lattice structure of a crystal. The comparison determines the lattice interval in terms of the known interval between the lines of the ruled grating. Thus we obtain the dimensions of the lattice in centimetres, and hence the number of lattice cells in unit volume, or in unit mass of the crystal. Since the lattice is formed by molecules, this gives the number of molecules in unit mass of the crystal. Using the known relative atomic weights, we deduce the mass M of a hydrogen atom in grams.

This appears to be the only practical way of determining M without introducing e or h. But we can make another determination *via* e. A theoretically simple way of finding e is to measure macroscopically the charge $2ne$ provided by a large number of α particles; the number n of particles, whose charges are collected, is ascertained by actually counting them. Further, e/M can be found by an electrolytic method—equivalent to accumulating the charges of a known mass of hydrogen until an amount large enough to measure macroscopically is obtained. By combining e and e/M, a value of M is obtained which is less than that found by the crystal grating method. The observed ratio agrees with $\frac{136}{137}$ within the limits of observational error.

There appears to be only one vulnerable link in this chain of connection, namely the assumption that in the crystal there is one molecule to a lattice cell. That depends on a classical picture of the particles in the crystal. In wave mechanics the structure of the crystal is represented by standing waves. It is not necessary to suppose that the "particles of composite identity", which are more or less located at the lattice points, represent an equal number of pure particles. In more elementary problems the invariance of the density of eigenfunctions secures that different modes of analysis into particles yield the same total number of particles; but this does not apply when the number of dimensions of the phase space is different in the two analyses. A crystal is a typical example of a system composed of indistinguishable elements. Any pair of elements will have a permutation coordinate and an interchange energy conjugate to it. It is possible, and even probable, that in constituting the standing waves of a system of this character, 137 particles of pure identity furnish 136 composite particles.

The problem lies rather beyond the point which our theory has reached. But if the observations are trustworthy, it would seem to be a direct deduction from them that there is a discrepancy in the two ways of counting particles in a macroscopic aggregation. In counting α particles, no question of interchange arises, because they are observed one at a time. Ideally we might vaporise the crystal, and count its molecules in the same way; but instead, we try to count the molecules as they lie simultaneously in space. It is scarcely surprising that we should become confused in our count of entities which have no definite position and are indistinguishable from one another. We count *something*; but not the particles which we should count if they passed successively before us. In the crystal we identify a molecule by the lattice cell which it occupies, disregarding the fact that the molecules counted by the other method could not be localised individually. This procedure introduces a permutation coordinate. The extra coordinate increases the energy invariant of a particle in the ratio $\frac{137}{136}$, in accordance with the law of equipartition of energy. We have thus to assign to each crystal molecule a mass $\frac{137}{136}$ times that of a corresponding gas molecule—as the observations actually indicate. This is not energy of crystallisation (which is too small to be considered); it is presumably a sign that, in the mathematical transformation of the external wave functions of gas molecules into internal wave functions of a crystal, 137 gas particles become repartitioned in 136 crystal particles. We can thus suggest a possible origin of the discrepancy of the crystal results; but it would require a much fuller investigation to determine whether the factor $\frac{137}{136}$ is an observable correction, or whether it is compensated in the complete theory of the experiment.†

† I understand that (in view of still more recent experiments) the present opinion is that the supposed crystal grating discrepancy is spurious.

CHAPTER XVI

THE EXCLUSION PRINCIPLE

16·1. The Second Quantisation.

Let J_r ($r=1, 2, \ldots n$) be n commuting idempotent symbols. Let U be a common eigensymbol of the J_r, and let e_r be the corresponding eigenvalue of J_r. Since $J_r^2-J_r=0$, $e_r=0$ or 1.

By § 3·7 (e) an eigensymbol for any given set of eigenvalues e_r can be found explicitly, namely

$$U=\left\{\prod_1^n (J_r-1+e_r)\right\} V. \tag{16·11}$$

where V is *any* symbol. Since $J_s^2-J_s=0$, and $e_s^2-e_s=0$,

$$(J_s-e_s)(J_s-1+e_s)=0, \tag{16·12}$$

from which it follows at once that $(J_s-e_s)U=0$ for all values of s.

Let
$$q=\prod_1^n (J_r-1+e_r). \tag{16·131}$$

We shall regard q as a transformation operator. It transforms any expression V_s, which is not in general an eigensymbol of the J_r, into

$$U_s=qV_s, \tag{16·132}$$

which is an eigensymbol giving the set of eigenvalues defined by q. By (16·12), q is singular; there is therefore no inverse transformation. Also by (16·12)

$$qJ_s=qe_s. \tag{16·133}$$

A case of special importance is

$$V=\sum_1^n J_sV_s, \tag{16·141}$$

which gives on transformation

$$U=\sum_1^n qJ_sV_s=\sum_1^n e_sU_s=\Sigma_e U_s, \tag{16·142}$$

where Σ_e denotes summation over those suffixes for which $e_s=1$.

Similarly, if
$$V=\Sigma J_rJ_sV_{r,s}, \tag{16·143}$$

we have
$$U=\Sigma e_re_sU_{r,s}=\Sigma_e U_{r,s}, \tag{16·144}$$

the summation Σ_e being restricted to pairs of suffixes both of which correspond to eigenvalue 1.

To apply this in physics, we take J_r to be an *existence operator* for an entity or condition ϵ_r. If J_r has the eigenvalue 1, ϵ_r is said to exist in the system described by the eigensymbol U; if the eigenvalue is 0, ϵ_r does not exist in U. In the system described by V which is not an eigensymbol of J_r, the entity

ϵ_r is not definitely present or absent, and ϵ_r is said to have partial existence in V. Two kinds of partial existence are commonly recognised; ϵ_r may have a probability of existing in V; or, without being wholly in V, it may have a component in V. In the former case, what may be called the "degree of existence" is expressed by a numerical relation to complete existence; in the latter it is expressed by a directional relation. Both are combined in a vector which possesses magnitude and direction; and, as we shall see later, the general existence symbol J_s, which replaces the classical existence symbols (1 for existence, and 0 for non-existence), can be represented as a strain vector.

To understand the importance of this investigation, we must recall that in the primary development of quantum theory a system is looked upon as the product of its parts rather than as the sum of its parts. We have treated double systems in this way in Chapter XII, and quadruple systems (a proton and electron with their respective comparison particles) in Chapter XV. But for large assemblages the product method becomes unsuitable. If the probability of one member of the assemblage vanishes, the probability of the whole assemblage vanishes. That takes us far from the macroscopic outlook in which the existence of one particle more or less scarcely matters. Now it is important to develop a microscopic theory which shall converge to the macroscopic theory when the number of particles becomes very great. The basis of this new development must be the assignment of additive instead of multiplicative characteristics to the parts into which the system is analysed.

To a certain extent we have anticipated this new development, especially in Chapters XIII and XIV. But we there freely employed the exclusion principle, and we now return to the beginning of the theory to discover its foundation.

In (16·142) we have practically the classical conception of an additive characteristic. The quantity U is made up of contributions U_s from the various particles ϵ_s, the summation being of course limited to the particles which exist in the assemblage to which U refers. The assemblage is,however, viewed from the standpoint of *selection* rather than *creation*; so that $\Sigma_e U_s$ appears as part of a universal sum ΣU_s. This is now treated as a particular case of an assemblage of partially existing particles with characteristics V_s, which are additive in a generalised sense $V = \Sigma J_s V_s$, the symbolic coefficient J_s being a measure of the degree to which the sth particle is present in the assemblage in the sense explained above. This generalised addition is of the same type as that used in expressing a vector as the sum of its components, $T = \Sigma E_\mu t_\mu$.

In calling the entities ϵ_r "particles" we are redefining the term *particle*, so as to correspond to the conception of a system as the sum of its particles

rather than as the product of its particles. As we have pointed out, the particles in macroscopic systems in equilibrium have composite identity as compared with the pure elementary particles originally introduced.

We shall refer to the present development as *secondary* quantum theory, the multiplicative combination of particles being *primary* quantum theory.

16·2. Jordan-Wigner Wave Functions.

We shall now express $V=\Sigma J_s V_s$ in a form, due to Jordan and Wigner, which is well known in quantum theory. This transformation is not of any particular importance from our point of view, $\Sigma J_s V_s$ being not only the most direct but also the most convenient expression of the additive combination rule; but it is desirable to show how it is connected with the customary outlook.

In the application contemplated, the suffixes s refer to the pure elementary states that can be occupied by particles, and V_s is the stream vector of the sth state. In general a state is only partially occupied; equivalently we can say that a particle whose stream vector is V_s is partially present in the system considered.

Since the states are pure, V_s is the product of two wave vectors. The Jordan-Wigner theory is limited to the case in which the stream vector of each state is a perfect square. We can therefore take V_s to be a strain vector of index 2 equal to $\psi_s.\psi_s$, or a strain vector of index 0 equal to $\bar{\psi}_s.\psi_s$, where $\bar{\psi}_s$ is the complex conjugate of ψ (§ 8·6). It is understood that these are outer products; it would be inconvenient in the present investigation to indicate this explicitly by the asterisk notation.

The characteristic equation $J_s^2-J_s=0$ of an idempotent operator is of the second degree, so that it can be represented by a twofold matrix. We shall therefore represent J_s by the idempotent twofold matrix

$$J_{s,\lambda}=\tfrac{1}{2}(1+i\zeta_\lambda), \tag{16·211}$$

where ζ_λ is a Pauli matrix, i.e. any degenerate twofold matrix whose square is -1. In treating $\Sigma J_{s,\lambda}V_s$, we wish to resolve $J_{s,\lambda}$ into two factors to be associated respectively with the two factors of V_s. It is possible to express $J_{s,\lambda}$ as the product of two 2-vectors; but in the Jordan-Wigner theory a different kind of factorisation is employed. Let ζ_λ, ζ_μ, ζ_ν be three perpendicular Pauli matrices. Then by (3·86)

$$\left.\begin{aligned}\tfrac{1}{4}(\zeta_\mu+i\zeta_\nu)(\zeta_\mu-i\zeta_\nu)&=-\tfrac{1}{2}(1+i\zeta_\lambda)=-J_{s,\lambda},\\ \tfrac{1}{4}(\zeta_\mu-i\zeta_\nu)(\zeta_\mu+i\zeta_\nu)&=-\tfrac{1}{2}(1-i\zeta_\lambda)=-(1-J_{s,\lambda}).\end{aligned}\right\} \tag{16·212}$$

We can therefore write

$$J_{s,\lambda}=\bar{a}_s a_s, \quad 1-J_{s,\lambda}=a_s\bar{a}_s, \tag{16·22}$$

where

$$\bar{a}_s=\tfrac{1}{2}i_s(\zeta_\mu+i\zeta_\nu)_s, \quad a_s=\tfrac{1}{2}i_s(\zeta_\mu-i\zeta_\nu)_s, \tag{16·23}$$

and i_s is a square root of -1, commuting with the ζ's.

Since the existence operators J_r, J_s commute, the Pauli matrices used for their representation commute; that is to say, we must employ for the different particles Pauli matrices $\zeta_{\lambda r}$, $\zeta_{\mu r}$, $\zeta_{\nu r}$ and $\zeta_{\lambda' s}$, $\zeta_{\mu' s}$, $\zeta_{\nu' s}$, belonging to different commuting symbolic frames.

The Jordan-Wigner wave functions are

$$\bar{\Psi} = \Sigma \bar{a}_s \bar{\psi}_s, \quad \Psi = \Sigma a_s \psi_s. \tag{16·24}$$

Forming their outer† products $\bar{\Psi}\Psi$ and $\Psi\bar{\Psi}$, we obtain, if we omit the cross-terms,

$$\left.\begin{aligned} \bar{\Psi}\Psi &= \Sigma \bar{a}_s a_s \bar{\psi}_s \psi_s = \Sigma J_s V_s = V, \\ \Psi\bar{\Psi} &= \Sigma a_s \bar{a}_s \psi_s \bar{\psi}_s = \Sigma (1 - J_s) V_s = \Lambda, \end{aligned}\right\} \tag{16·25}$$

by (16·22). We call Λ the *complementary stream vector* to V. It represents the stream vector of the states to the extent to which they are unoccupied, in the same way that V is the stream vector of the states to the extent to which they are occupied.

The omission of the cross-terms is to some extent justified when the elementary wave functions ψ_s form an orthogonal set, so that their products vanish on integration over the whole domain. It is in this integrated sense that $\bar{\Psi}\Psi$ and V are equivalent. Physically this correspondence is sufficient, because we have no right to treat a wave function except as a whole. But, if we consider the universe of completely occupied states, given by $\Sigma V_s = V + \Lambda$, we come back to the classical conception of addition of V_s. In order that our formulae may converge to those of classical theory, it is necessary that $\bar{\Psi}\Psi + \Psi\bar{\Psi}$ should converge to $V + \Lambda$ locally as well as on integration; that is to say, the cross-terms in $\bar{\Psi}\Psi + \Psi\bar{\Psi}$ must vanish.

The point is that when a state extending over a region of space is partially filled, the location of the entity which is (partially) occupying it is a matter of probability; and we have the complication usual in statistical theory that the expectation value of a product is not the same as the product of the expectation values. But when the state is fully occupied, it represents a definite distribution which can be treated as a classical fluid having a determinate density at every point.

The rs cross-term in $\bar{\Psi}\Psi + \Psi\bar{\Psi}$ is

$$\bar{a}_r a_s \bar{\psi}_r \psi_s + \bar{a}_s a_r \bar{\psi}_s \psi_r + a_r \bar{a}_s \psi_r \bar{\psi}_s + a_s \bar{a}_r \psi_s \bar{\psi}_r.$$

Since the ψ's commute, this will vanish if

$$\bar{a}_r a_s + a_s \bar{a}_r = 0, \quad \bar{a}_s a_r + a_r \bar{a}_s = 0. \tag{16·261}$$

Since $\zeta_{\mu r}$ and $\zeta_{\mu s}$ commute, (16·261) will be satisfied if

$$i_r i_s + i_s i_r = 0. \tag{16·262}$$

† Outer products so far as ψ_s and $\bar{\psi}_s$ are concerned. The symbolic coefficients follow their own commutation rules found below; ψ_s and $\bar{\psi}_s$ commute.

That is to say, the factors i_s in the different terms must be taken to be anticommuting square roots of -1. Hence we have also

$$\bar{a}_r\bar{a}_s+\bar{a}_s\bar{a}_r=0, \quad a_ra_s+a_sa_r=0 \qquad (r\neq s). \tag{16·271}$$

Also by (16·22) $\bar{a}_sa_s+a_s\bar{a}_s=1$. Combining this with (16·261), we obtain

$$\bar{a}_ra_s+a_s\bar{a}_r=\delta_r{}^s. \tag{16·272}$$

The equations (16·271) and (16·272) satisfied by the coefficients of the wave functions (16·24) are known as the Jordan-Wigner commutation rules.

16·3. Einstein-Bose Particles.

In the foregoing analysis we have supposed that the states or particles are fully distinguished by the additive characteristic V_s which is being studied. We shall now consider states with a distinguishing characteristic W_{st}, such that the states $W_{s1}, W_{s2}, \ldots W_{sm}$ all have the same V_s. Let J_{st} be the existence operator for the particle occupying the state W_{st}; then (16·141) is replaced by

$$V=\Sigma_{s,t}J_{st}V_s=\Sigma K_sV_s, \tag{16·311}$$

where

$$K_s=J_{s1}+J_{s2}+\ldots+J_{sm}. \tag{16·312}$$

Since the J_{st} commute and have eigenvalues 0, 1, the eigenvalues of K_s will be 0, 1, 2, ... m. Its minimum equation is therefore

$$K(K-1)(K-2)\ldots(K-m)=0. \tag{16·32}$$

Let

$$K'=K-\tfrac{1}{2}m.$$

Then, if m is even, (16·32) gives

$$K'\prod_{-\frac{1}{2}m}^{\frac{1}{2}m}\left(1+\frac{K'}{r}\right)=0, \tag{16·33}$$

where r takes successive integral values from $-\frac{1}{2}m$ to $\frac{1}{2}m$. As $m\to\infty$, this tends to the limit

$$\sin\pi K'=0. \tag{16·34}$$

Similarly, if m is odd, the limit is $\cos\pi K'=0$.

Particles, of which there may be any number in the same recognised state V_s, are called *Einstein-Bose particles*, as distinguished from excluding or *Fermi-Dirac particles*. The distinction is practical rather than fundamental; the Einstein-Bose case arises when the *recognised* states V_s comprise a great number of ultimate elementary states W_{st}, i.e. when the practical classification of states is less minute than the theoretical classification. We have then to associate a *cardinal operator* K_s instead of an existence operator J_s with the recognised state, and its eigenvalue gives the number of particles in that state.

The use of K' instead of K depends on the theory of the self-consistent field. We cannot make a large change in the number of particles in a state without upsetting the conditions which governed the original analysis into

states. Strictly speaking we must know what states are occupied in order to determine the controlling field, and a knowledge of the controlling field must precede the analysis into states. In practice this means that we employ an approximate or average distribution of the particles to calculate the self-consistent field. Then we must provide symmetrically for the addition of extra particles, and the subtraction of particles allowed for but not present. The characteristic to be considered is therefore $V' = \Sigma K_s' V_s$, where $2K'$ is an operator which takes the eigenvalues 1, 2, 3, ... according to the number of additional particles definitely present, and $-1, -2, -3, \ldots$ according to the defect in the number of particles. This is fulfilled by operators satisfying the characteristic equation $\sin \pi K' = 0$.

The substitution of K_s' for K_s is equivalent to the substitution $J_{st} - \frac{1}{2}$ for J_{st} in (16·311). We may regard half of the energy (or other mechanical characteristic) of the particle as allotted to the self-consistent field and thereby contributing to the energies of the particles in other states. Then if the particle is fully present we have to add its half of the mutual energy; if it is absent we have to subtract the energy wrongly included in the self-consistent field. This is provided for by the eigenvalues $\frac{1}{2}, -\frac{1}{2}$ of $J - \frac{1}{2}$.

Another point of view is that $(J - \frac{1}{2})$ and K' are *relative* existence operators, as distinguished from the *absolute* existence operators J, K. It would be absurd to schedule everything which does not exist. To call attention to the non-existence of an entity implies that we had some *a priori* expectation that it would exist. To regularise this outlook we contemplate a comparison system consisting of entities ϵ_r which have an equal probability of existing or not, so that their respective existence operators have expectation values $\frac{1}{2}$. The object-system is then specified by relative operators $J_r - \frac{1}{2}$, which express deviation from the comparison distribution. This procedure in secondary quantum theory corresponds to that which we have already introduced in primary quantum theory in § 11·6. In connecting the results of primary and secondary theory it is to be noticed that the ideal comparison fluid of secondary theory consists of a set of *half-occupied* states, whereas the comparison fluid of primary theory is the impermeable background of *fully-occupied* states formed by the unspecified particles of the universe.

In the usual applications the number m is treated as infinite, so that (16·34) applies. A solution is

$$K_s' = -i\partial/\partial\theta_s, \tag{16·35}$$

where θ_s is a periodic argument of all operands contemplated. For then

$$e^{i\pi K_s'} f(\theta_s) = e^{\pi\partial/\partial\theta_s} f(\theta_s) = f(\theta_s + \pi) = f(\theta_s - \pi).$$

Thus $e^{i\pi K_s'} = e^{-i\pi K_s'}$; so that $\sin \pi K_s' = 0$. We thus obtain a representation of V' as

$$V' = \Sigma K_s' V_s = -i\Sigma \partial V_s/\partial\theta_s. \tag{16·361}$$

It is appropriate that a new coordinate θ_s should be introduced, since the existence of states W_{st} not discriminated by the values of V_s implies that there exist other characteristics, e.g. momenta $\partial/\partial\theta_s$, unrecognised. A connection with classical theory is obtained by setting

$$\theta_s = 2\pi\nu_s t,$$

where t is a common argument for all the functions V_s. Then ν_s is the frequency of V_s; but V_s is not necessarily a simple harmonic function. By (16·361)

$$V' = -\frac{i}{2\pi}\Sigma\frac{1}{\nu_s}\frac{\partial V_s}{\partial t}. \tag{16·362}$$

In particular, if V_s is a momentum vector $-i\partial/\partial x_\mu$, the resultant momentum vector is

$$V' = -\Sigma\frac{1}{2\pi\nu_s}\left(\frac{\partial^2}{\partial x_\mu \partial t}\right)_s \tag{16·363}$$

which may be compared with the momentum $\mathbf{T}_{\mu 4}$ given by the energy operator (13·12).

In current quantum theory Einstein-Bose particles are represented by wave functions analogous to the Jordan-Wigner wave functions of Fermi-Dirac particles. It is assumed as before that the stream vector V_s is a perfect square. V' is written in the modified form

$$V' = \Sigma\bar{\psi}_s K_s'\psi_s = \Sigma\bar{\psi}_s\bar{a}_s a_s\psi_s.$$

Since K_s' now operates on one factor only, we must take $K_s' = -2i\partial/\partial\theta_s$.

A possible factorisation of K_s' is

$$\bar{a}_s = e^{\frac{1}{2}i\theta_s}, \quad a_s = \partial/\partial\bar{a}_s, \tag{16·371}$$

since $\bar{a}_s a_s = \partial/\partial(\log\bar{a}_s) = -2i\partial/\partial\theta_s$. Thus, if cross-terms are omitted,

$$V' = \bar{\Psi}\Psi = (\Sigma\bar{\psi}_s\bar{a}_s)(\Sigma a_s\psi_s). \tag{16·372}$$

Further, if f is any symbol,

$$\frac{\partial}{\partial\bar{a}_s}(\bar{a}_s f) - \bar{a}_s\frac{\partial}{\partial\bar{a}_s}f = f,$$

so that $a_s\bar{a}_s - \bar{a}_s a_s = 1$. And since all other combinations of a_s, $\bar{a}_s$ commute, the complete commutation rules are

$$a_s\bar{a}_r - \bar{a}_r a_s = \delta_r{}^s, \tag{16·381}$$

$$\bar{a}_s\bar{a}_r - \bar{a}_r\bar{a}_s = 0, \quad a_s a_r - a_r a_s = 0. \tag{16·382}$$

These may be compared with the Jordan-Wigner commutation rules in (16·271), (16·272).

In problems concerning a specified system, the unspecified particles will in general be Einstein-Bose particles. For it is clear that the type of analysis into states V_s applied to the specified system does not provide anything like enough states to accommodate the 10^{79} unspecified particles separately.

For example, consider the box-problem (§ 13·2) applied to material of the density of water. If (to avoid introducing supernatural barriers) we suppose the uniform density to continue indefinitely, we obtain a space whose radius is of the order 3.10^8 km., and a mass about 10^{-14} of the mass of the universe. In this case the recognised states provide room for only 1 in 10^{14} particles; so that in (16·312) $m = 10^{14}$. The unspecified particles provide the impermeable background with occasional vacancies which we have identified with photons. Any number of vacancies up to 10^{14} may occur in one recognised energy state; and the photons accordingly obey Einstein-Bose statistics.

It will be seen that a change of outlook can have the effect of changing Fermi-Dirac particles into Einstein-Bose particles. The theory of Chapter XIV is based on a system of excluding particles (system A); but these become transformed into Einstein-Bose particles in the usual outlook (system B).

16·4. Relation to the Energy Tensor.

It is important to realise the considerable change of outlook that is involved in passing from the primary to the secondary quantum theory. The analysis which shows that in given field conditions there will be a series of steady states, composed of probability distributions with wave functions ψ_s or stream vector functions V_s, is common to both. In primary quantum theory a particle may have its total probability (unity) distributed between different states, so that its stream vector is $V = \Sigma p_s V_s$. If a second particle is present with the stream vector $V' = \Sigma p_s' V_s$, there is no warrant for forming the sum $\Sigma (p_s + p_s') V_s$. The only combination of V and V' contemplated is the product VV', which indicates a set of double states $V_r V_s'$ occupied with probabilities $p_r p_s'$. If the addition $p_s + p_s'$ were admitted, the total probability of a state might exceed unity, and the exclusion principle would be violated; this contingency does not arise, since the conception of such an addition is foreign to the outlook of primary quantum theory. In secondary quantum theory a different type of particle is introduced (composite from the point of view of primary theory), each particle being limited to one state. The numerical probability factor p_s is now replaced by a vector probability factor J_s. The exclusion principle is represented by the fact that, the eigenvalues of J_s being 0 and 1, its expectation value will always lie between these limits.

The general theory of relativity throws light on the meaning of $V = \Sigma J_s V_s$. Macroscopically the additive characteristic of a system is its energy tensor. Thus the secondary quantum theory rightly rejects the sum of vectors ΣV_s or $\Sigma p_s V_s$, and employs the sum of tensors of the second rank $\Sigma J_s V_s$. Moreover, owing to the non-linearity of the equations, the self-energy tensors of particles are not strictly additive in general relativity theory. The additive energy is a mutual energy of the particle and the rest of the universe. The

energy is therefore not the square of the momentum vector $V_s V_s$, but the product of a special factor V_s and a world factor J_s. We have already seen how we endeavour to reconcile this with our usual non-relativistic outlook by artificially breaking up the mutual energy into self energies of the particle and the rest of the universe (treated as comparison fluid). The convergence of the energy tensor $\Sigma J_s V_s$ of secondary quantum theory to the energy tensor $T_{\mu\nu}$ of general relativity theory, as we pass from the microscopic to the macroscopic outlook, is exhibited in (16·363). Since the macroscopic outlook implies a much less minute discrimination of individual states, the quantum expression for the energy tensor is first reduced to Einstein-Bose form $\Sigma K_s' V_s$.

16·5. Enumeration of Wave Functions.

By using the observed values of e/m and κ, we have found that there are approximately 136.2^{256} protons and an equal number of electrons in the universe. This will now be confirmed by an independent investigation which shows that the number is exact.

Our ability to predict the number of protons and electrons in the universe implies that the number is imposed by the procedure followed in analysing the interrelatedness of our experience into a manifestation of an assemblage of particles or wave systems. It is a commonplace that electrons are not intrinsically distinguishable from one another; it is therefore not surprising that the total number, allowed for in our scheme of dissection of phenomena, depends on the conventional distinctions introduced when, for example, we decide that a certain diffuse wave packet is composed of two electrons rather than one.

We first notice that the value we have found for the total number of particles N belongs to a series of numbers

$$N_1 = 2.3.2^4, \quad N_2 = 2.10.2^{16}, \quad N_4 = 2.136.2^{256},$$

i.e. $n(n+1)2^{n^2}$, with $n = 2, 4, 16, \ldots$. If N_1 is regarded as characteristic of a simple wave function, N_2 and N_4 will be the corresponding characteristics of double and quadruple wave functions. We shall first investigate a "universe" described by simple wave functions and show that it consists of N_1 (i.e. 96) particles. It will then be comparatively easy to show that the actual universe contains N_4 particles.

Conditions—we can scarcely call them phenomena—which require for their representation no more than a simple matrix frame have been treated in Chapters I to VIII. We can represent a position vector or a stream vector, but not both; if the position vector is represented in an E-frame, we have to change to a C-frame to represent the stream vector (§ 10·5). The familiar particle, whose position vector and stream vector are both inexact but not

entirely uncertain, cannot be represented on this plan. But in Chapter VIII we made a new departure by introducing D-operators by means of which we were able to define vector functions instead of isolated vectors; and in particular the momentum vector was defined. In fundamental investigations of the kind which we are now treating, the stream vector and the momentum vector coalesce;† but their coalescence is, strictly speaking,

$$E_\mu(-i\,\partial/\partial x_\mu) = E_\mu p_\mu = C_\mu j_\mu. \qquad (16{\cdot}51)$$

For the operator $-i\partial/\partial x_\mu$ applies to a representation of positions, and it is necessary to change to another frame to obtain a representation of the stream vector j_μ.

Consequently, by using position and momentum instead of position and stream, we are able to keep the whole representation of the particle within one matrix frame—at the cost of using differential operators as well as the matrix operators of the frame.

This is an illustration of the use of differential operators to avoid multiplicity of frames. It can be greatly extended; and, as is well known, particles are generally distinguished by allotting to each a different orthogonal wave function, rather than by allotting a different matrix frame as contemplated in § 2·9. It is true that we must still assign to them existence operators J_s which are represented by Pauli matrices in different frames; but that, as it were, segregates the question of multiplicity of frame from the main part of the analysis, and eliminates it altogether in the case of particles having complete existence in the system ($J_s = 1$).

In the systems containing many particles which we have hitherto treated —the box problem in Chapter XIII and the Einstein universe in Chapter XIV —we were concerned only with fully occupied states, so that existence operators (other than unity) with their attendant multiplicity of frames did not appear. If the scalar state is not fully occupied, we must in general express the extent of its occupation by an idempotent operator J. In (13·61) we divided the scalar state into two substates J and $1-J$, with equal and opposite spin. The purpose of dividing an entity into two parts is that we may conceive of one part existing without the other; so that the division is made by distinguishing a part J which is occupied (since $J.J = 1.J$) from a part $1-J$ which is unoccupied (since $J(1-J) = 0$). The further separation of the state into substates with opposite charges depends on a double existence operator $J_r J_s'$ for two entities, since charge is meaningless unless there are two charges to interact. The four substates (13·62) are represented by $J_r J_s'$, $(1-J_r)J_s'$, $J_r(1-J_s')$, $(1-J_r)(1-J_s')$.

† We regard non-algebraic wave functions, whose momentum vector has matrix components and is not to be identified with j_μ, as a side-development of the wave functions primarily studied. As explained in § 9·5, the steady states defined by them are only conditionally steady.

The system represented by a simple wave function has four dynamical coordinates, and in field-free conditions we may take the steady state to be (cf. (8·53))

$$\psi(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = e^{\frac{1}{2}i(\pm E_{12}\alpha_1 \pm E_{23}\alpha_2 \pm E_{31}\alpha_3 \pm E_{14}\alpha_4)}\psi_0, \qquad (16\cdot52)$$

where ψ_0 is a common eigensymbol of the four commuting E-symbols. The dynamical angular momenta $-i\partial/\partial\alpha_1$, etc., have eigenvalues $\pm\frac{1}{2}$. Each set of four values defines a corresponding eigenfunction ψ, and the 16 combinations of sign therefore define 16 eigenstates.

In our earlier treatment we regarded a particle as distributed with probability factors p_r between the different eigenstates; to introduce a second particle would involve beginning the problem over again with double wave vectors. In secondary quantum theory the states are considered to be occupied by 16 different partially existing particles, which in special cases may be definitely present or absent. We have therefore to assign to them existence operators J_r ($r = 1, 2, \ldots 16$) stating the degree of occupation.

The expression (16·52) singles out a particular plane E_{23} as the plane of spin, and this choice will be reflected in the value of the eigenvector ψ_0. The choice might equally well have been E_{31} or E_{12}. This will be investigated in greater detail in the next section; but it is fairly evident that the effect of these alternatives will be to triplicate the set of independent wave functions, making the total number 48. This number is again doubled when we allow for two possible algebraic frames $(1, \pm i)$, or an equivalent duplexity. The total then becomes 96 or N_1.

We now begin to see the make-up of the numbers N_1, N_2, N_4. The elementary states correspond to wave functions of the form $e^{\frac{1}{2}i\Sigma(\pm i\alpha_r)}\psi_0$. The factor 2^{n^2} corresponds to the variety of the exponential; and the factor $n(n+1)/2$ corresponds to the variety of ψ_0; finally there is a factor 2 corresponding to the duplexity of the algebraic frame. We notice that in N_1, the four dynamical coordinates give a factor 2^4, whereas the three geometrical coordinates give a factor 3. This is because the wave function depends on a *combination* of eigenvalues of the dynamical momenta and a *choice* of geometrical axis. Similarly in N_2 and N_4 the large factor arises from the number of combinations of eigenvalues of operators with a common eigensymbol, and the small factor represents a choice of "orientation" of this eigensymbol.

The final doubling of the number is the most difficult step to investigate. It is fairly obvious that the duplexity of the algebraic frame will contribute a factor 2; but it is less easy to be sure that the factor has not already been included in the enumeration.

Before studying the method of enumeration more rigorously, we must make clear the connection between the existence of N_1 eigenfunctions of

type (16·52) and the formation of a corresponding universe of N_1 particles. In Chapter XIV the particles constituting the Einstein universe were represented by wave functions in a *flat* space, formed by stereographic projection of the "actual" spherical space. Here we begin at the other end of the problem; for the wave functions (16·52) are those occurring in elementary quantum theory, which, as we have seen (§ 14·3), presupposes an Einstein universe as background. These wave functions occupy actual space; and the proper energy of the N_1 particles, being represented by the curvature of space, cannot be represented a second time in the wave functions themselves. By the dynamical theory these eigenfunctions give the only possible steady states of disturbance of the system; there are therefore just N_1 independent steady modifications (generalised states of rotation). But, owing to the symmetry of the conditions, there is a degeneracy which permits us to re-analyse their most general combination into various alternative sets of N_1 elementary states; in particular we can analyse it into an equivalent number of spherical harmonic distributions. By projecting these into a flat space, we restore to the wave functions the proper energy which was abstracted to provide the curvature, and obtain wave functions corresponding to those of Chapter XIV. The latter, by definition, correspond to steady states in the self-consistent field produced by the whole aggregation, so that they must conform to this method of enumeration.

Lemaître has pointed out to me that the ambiguity $\pm E_{16}\alpha_4$ in (16·52) seems to be in conflict with (8·54), where the algebraic dynamical coordinate was given unique sign. But this ambiguity is seen to be necessary, when we remember that the set of wave functions, when fully occupied, must not provide a resultant proper energy additional to that represented by the curvature of the space to which they are referred. In primary quantum theory reversal of the algebraic coordinate would turn the particle into a minus-particle; and here the double sign will imply that in some sense the 96 composite particles can be regarded as 48 plus-particles and 48 minus-particles—though we must not too hastily assume that the latter are positrons and negatrons. Since we are treating a set of particles whose resultant energy tensor provides the curvature of the space in which their individual wave functions are represented, the contributions of these wave functions to the energy tensor must cancel out, and there must be as many minus-particles as plus-particles in the set. It is therefore right to include the minus-particles in the enumeration. This relative representation provides the simplest method of counting the independent wave functions; but to obtain the particles ordinarily recognised in physics we must restore to them the energy abstracted to form the curvature. By this transformation they all become plus-particles,

as in § 14·2. To put the conclusion in another form: elementary quantum theory always describes its systems as additions to the impermeable background which constitutes the Einstein universe. If therefore we describe by the wave functions of elementary quantum theory the set of particles which composes the impermeable background, the plus-additions and minus-additions must balance. But the minus aspect is only relative, and the particles are all positive contributors to the total energy.

16·6. Geometrical Representation of J_r.

By squaring (16·52) we obtain the strain vector of index 2

$$V(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = e^{\pm i\alpha_1 \pm i\alpha_2 \pm i\alpha_3 \pm i\alpha_4} V_0, \qquad (16·611)$$

the matrices being eliminated because V_0 is an eigensymbol. Since $V_0 = \psi_0\psi_0$, it has the pure form

$$V_0 = \tfrac{1}{4}(1 + i\zeta_u)(1 + i\theta_v). \qquad (16·612)$$

We can distinguish the possible functions V as $V_r(u, v)$, the suffix r ($r = 1, 2, \ldots 16$) indicating a particular combination of signs in (16·611) and the arguments u, v indicating planes in the two Pauli frames.

Consider the existence operator J of V. J is represented in a Pauli frame; and since there is a possibility that V may be an eigensymbol of J, i.e. that $JV = V$, this must be one of the two frames of V. Thus the ζ and θ frames have the fundamental distinction that one of them (namely the spin frame ζ) is also the frame of the existence operator. To interpret this physically we recall that J is the world factor in the energy tensor JV (§ 16·4). In the conditions represented by a simple wave tensor the "rest of the universe" is understood to form a uniform static distribution defining an electrically neutral three-dimensional space. On the other hand V involves four dimensions and is not electrically neutral; thus it contains an additional factor $(1 + i\theta_v)$ whose plane v is not definable by reference to the three-dimensional space of J.

Let $J_w = \frac{1}{2}(1 + i\zeta_w)$ be the existence operator of a particle ϵ_w. By (16·612) $J_w V = V$ if $\zeta_u = \zeta_w$, and $J_w V = 0$ if $\zeta_u = -\zeta_w$. What we have hitherto called the degree of existence of the particle ϵ_w in the system V, is reduced to a directional relation between the spin planes ζ_w, ζ_u of ϵ_w and V. This simplification is due to the fact that V is a *pure* strain vector, and could itself be regarded as representing an elementary particle ϵ_u with an existence operator J_u. In our original formulae (§ 16·1) V was not restricted to be a pure strain vector; it might therefore have a constituent agreeing with ϵ_w in direction but reduced in intensity by a numerical probability factor. This mode of partial existence of ϵ_w in V is now excluded by the condition of purity. Accordingly the degree of failure of one elementary particle or wave system to exist fully in another elementary wave system is represented exclusively by angular deviation, and not by fractional probability.

To enumerate the independent eigenfunctions, we must replace $V_r(u, v)$, which has a continuously varying parameter u, by an appropriate number of functions $V_r(u_s, v)$ with fixed parameters u_s.† The continuity of u is a degeneracy due to the symmetrical conditions postulated, which permit relativity rotation in three dimensions (cf. § 13·6). To obtain a corresponding non-degenerate problem, we should limit the spin to three coordinate planes. Since the spin can be in either direction in each plane, this gives six alternatives. We conclude therefore that the 16 eigenfunctions $V_r(u, v)$ have a sixfold degeneracy, and correspond to 96 non-degenerate eigenfunctions.

The conclusion that the degeneracy is sixfold, not threefold, can be reached in another way. Pauli space is not an ordinary space of three dimensions. The Pauli matrices ζ_λ, ζ_μ, ζ_ν correspond to E_{23}, E_{12}, E_{31}, which have a different group relation from the matrices E_1, E_2, E_3 which represent rectangular axes in ordinary space. Suppose that rectangular axes Ox, Oy, Oz are laid down geometrically in a three-dimensional space. Taking it to be an ordinary space, we associate matrices E_1, E_2, E_3 with Ox, Oy, Oz, respectively; in so doing we do not give any chiral quality to the space. Taking it to be a Pauli space, we associate Pauli matrices ζ_λ, ζ_μ, ζ_ν with Ox, Oy, Oz, respectively; these satisfy $\zeta_\mu\zeta_\nu = \zeta_\lambda = -\zeta_\nu\zeta_\mu$ (3·86). But we can also associate with the same geometrical axes an alternative (left-handed) Pauli space by using the matrices in the order ζ_λ, ζ_ν, ζ_μ. Treating ζ_u as a vector referred to rectangular axes, $V_r(u, v)$ will have a threefold degeneracy; but there will also be a duplicate set of 48 eigenfunctions whose existence operators (or spins) are represented in the alternative Pauli space.

It might be suggested that all possible reversals have already been provided for by the $\pm$ signs in (16·52). But a consideration of the hydrogen atom makes it clear that the reversal of spin is additional. The orbital motion of an electron in any plane may be in either direction. In addition, two states are discriminated according as the electron spin is with or against the orbital motion. From this aspect the factor 2 arises from the duplexity indicated by the "fourth quantum number".

16·7. Double and Quadruple Wave Functions.

The enumeration of independent double wave functions proceeds on the same lines as the enumeration of simple wave functions. The maximum number of commuting symbols $E_\mu F_\nu$ is 16, obtained by combining an antitetrad of E-symbols with an antitetrad of F-symbols. We have therefore 16

† The direction of v in its 3-space does not matter. There is no reference standard to compare it with; and, whatever the direction may be, it is called the time. (It should be borne in mind that the physical ideas illustrated by a universe composed of simple wave functions are necessarily somewhat crude.)

dynamical coordinates α_μ, and the double strain vector function for a steady state, corresponding to (16·611), is

$$V(\alpha_\mu) = e^{\frac{1}{2}(\pm i\alpha_\mu)} V_0,$$

where V_0 is a common eigensymbol of the 16 commuting EF-symbols. The sixteenfold ambiguity $\pm$ provides 2^{16} combinations.

First suppose that V_0 is the product of two simple strain vectors. Then in place of (16·612) we have

$$V_0 = \tfrac{1}{16}(1+i\zeta_u)(1+i\theta_v)(1+i\zeta_{u'})(1+i\theta_{v'}).$$

An existence operator for a system of two particles is of the form

$$K = JJ' = \tfrac{1}{4}(1+i\zeta_w)(1+i\zeta_{w'}),$$

since the system does not exist unless both the particles exist. The two Pauli frames ζ, ζ' will combine into a frame of fourfold matrices E_μ', and the frames θ, θ' into another frame F_μ'.† Then K is a simple strain vector in the E'-frame and V_0 is a double strain vector in the $E'F'$-frame.

The strain vector V_0 is by definition a perfect square $\Psi_0\Psi_0$, but in general Ψ_0 is not the product of two simple wave vectors. The system represented by a double wave tensor is a superposition with different probabilities of a number of resolvable double systems. The essential result for our purposes is that the geometrical relations, which replace the conception of partial existence, are contained entirely in the E'-frame. Rotation in the F'-frame can be ignored, since there is no reference standard in that frame.

We have then a series of wave functions $V_r([uu'],[vv'])$, $r = 1, 2, \ldots 2^{16}$, involving a continuous parameter $[uu']$; and we have to determine the degree of degeneracy implied by the continuous parameter. If V_r represented two completely separate particles, $[uu']$ would represent their planes of spin in three dimensions. Since the particles are not separate $[uu']$ is a simple strain vector in an ordinary fourfold frame. This brings us to familiar ground. Although the domain in which we picture this vector is not ordinary space-time, the problem is the same mathematically. The relativity rotations, whose existence brings about the degeneracy, are the 10 rotations in five dimensions or the ten displacements in phase space, which we have studied fully in the earlier chapters. There is accordingly a tenfold degeneracy; or, including the duplication due to right- and left-handed frames of E_μ', each V_r is equivalent to 20 non-degenerate wave functions. This gives the total $N_2 = 2.10.2^{16}$.

It is perhaps not obvious that the number of rotations to be taken into account is 10 not 16. We might appeal to the formal argument that the corresponding factors in N_1 and N_2 should be either 3, 10 or 4, 16; and since the factor in N_1 is certainly 3, the factor in N_2 must be 10. But the result is

† This regrouping has a resemblance to the crossing of frames in § 10·4; but it is not the same transformation.

verified at once, when we notice that in comparing our double system with a standard system of two particles defined by the existence operator K, we cannot distinguish between the combination $J_w J_{w'}$ and $J_{w'} J_w$. Thus our reference system does not distinguish between a rotation $\zeta_\mu \zeta_\nu'$ and $\zeta_\nu \zeta_\mu'$, and the rotations $\zeta_\mu \zeta_\nu' - \zeta_\nu \zeta_\mu'$ must be ignored in the same way that we ignore rotations in the $\theta\theta'$ frame. Adapting the theory of crossed frames in § 10·6 to twofold matrices, the excluded rotations are the antisymmetrical (time-like) rotations in the crossed frame; there remain the ten symmetrical (space-like) rotations.

The extension to quadruple wave functions does not raise any new point. The reference system of four separate particles defined by a quadruple existence operator has a 136-dimensional phase space, so that the degeneracy is 136-fold.

In each case the degeneracy of the wave function corresponds to the phase space of a tensor of half its rank. A corresponding reduction occurs in the theory of the steady internal states of the hydrogen atom, which can quite well be (and commonly are) treated by Pauli vectors instead of Dirac vectors; but any deviation from the steady state involves the use of Dirac vectors.

16·8. Four-point Elements of Structure.

We can *observe* a relation between two physical entities. To *measure* a relation we must compare it with another relation of the same kind. Thus a measure is a relation between two relations and involves four entities.† Formally a measurement always refers to just four relata, viz. the terminals of the object relation and the comparison relation. But in practice systems of combination of measurements have been elaborated which enable us to attribute measure indirectly to more complex networks of relationship.

The basis of measurement is therefore a four-point element of world structure. It is on this principle that I have developed the generalised field theory in *Mathematical Theory of Relativity*.‡ Ultimately the theory of atomicity springs from the same origin.

In the field theory it was necessary to couple with the four-point relation a condition of affine geometry, viz. that for infinitesimal relations, if the relation AB is equivalent to the relation CD, then the relation AC is equivalent to the relation BD. More light is now thrown on this axiom by the transformation theory. I think that in attempting to compare or describe the difference between two objects, we are impelled to envisage a chain of intermediate objects—to see, as it were, one gradually changing into another. If such a transformation is too far-fetched to be considered, we

† In special cases two of the entities may coalesce.

‡ Chapter VII, Pt. II, especially § 98.

are nonplussed, as when asked (in the familiar riddle) the difference between an orange and a grand piano. Whether or not this is true of all relations, we do in physics confine attention to relations which are represented by transformations. The axiom of affine geometry then reduces to the condition that infinitesimal transformations commute as far as terms of the first order.

From this starting point we deduce the existence of a measure of world structure at each point, given by an affine curvature tensor $*B_{\mu\nu\sigma}{}^{\epsilon}$, which on contraction yields a metrical tensor $g_{\mu\nu}$ and an electromagnetic tensor $F_{\mu\nu}$. The tensor $g_{\mu\nu}$ provides a Riemannian geometry, and with this we define a new indirect measure system. A measure is still a relation between relations; but we now distinguish sharply between the object relation and the comparison relation. The latter is standardised; and it is no longer a relation between specific entities, but is vaguely "contained" in the Riemannian geometry. The measure is transferred verbally (and usually also in conception) to the object relation, so that it appears to involve only two entities. Finally in the non-relativistic mode of thought which dominates current quantum theory (whether styling itself relativistic or otherwise) the measure becomes transferred to one terminal of the object relation, a geometrical concept (e.g. an origin of coordinates) being substituted for the other terminal. By this devious route we arrive at entities supposed to be endowed with measurable properties, e.g. an electron endowed with charge and mass, although it requires four entities to furnish anything measurable.

In wave mechanics a distinction is drawn between observables and unobservables. It will here be clearer to call them measurables and unmeasurables. Current wave mechanics attributes measurables (momenta, coordinates, spins, etc.) to single entities such as an electron. The measurables are, however, four-point relations, which become attached to the electron because the other three relata are standardised. In the method of wave mechanics measurables are expectation values or eigenvalues derived from wave functions; and in attaching the measurables to the electron we attach also the wave functions which contain them. Primitively the wave function of an electron is that of a four-point element of structure.

The wave function is quadruple, since it specifies a quadruple probability distribution of four entities. We have found that the number of independent quadruple eigenfunctions is $N_4 = 2.136.2^{256}$. It is well known that (with due allowance for degeneracy when it occurs) the number of eigenfunctions persists in all transformations. By the changes in plan of measurement, to which we have referred, the quadruple wave functions become replaced by double and finally by single wave functions. Our position has been that, for better or worse, current theory has made this substitution; and since it is implied in all standard nomenclature, we must accept it. We have seen that certain previously inexplicable features in current physics, such as the mass-

ratio of the proton and electron, are impositions, which we are forced to accept in order to validate the substitution. Another such result is that by substituting N_4 simple wave functions for the N_4 quadruple wave functions, $2.136.2^{256}$ simple wave functions are forced into a space-time which is built to accommodate 96. The result of this overcrowding is that instead of their wave lengths being naturally adjusted to the curvature of space, they occupy spherical harmonics of high order and short wave length as explained in Chapter XIV.

Our result, that the "number of particles in the universe" is the number of independent non-degenerate eigenfunctions in a quadruple wave system, thus comes from the fact that a measurement, or comparison of relation with relation, is an expectation value determined by a quadruple wave system.

The number of dimensions of space-time can be regarded as one of the numerical constants of nature. Effectively, the number 4 was assumed at the beginning of our theory, when we chose a symbolic frame based on 4 anticommuting square roots of -1. But we can now determine it in the same way as the other natural constants, and show a fundamental reason why space-time has 4 dimensions and no more.

Space-time may be defined as the continuum in which we represent a certain type of physical relation called "displacement". The relation connects two relata which are conceived geometrically as points "having no parts and no magnitude". We need not add that they have position. Their *relative* position is expressed by the relation of displacement already mentioned; *absolute* position is denied. To pass from this geometrical abstraction to a physical universe, we must turn the points into physical entities; so that "displacement" is now a relation between entities occupying the points. These entities are represented by simple (commuting) existence operators, which indicate whether in a given operand (representing a state of the universe) the points are occupied or not. Nothing more is required; for, so far as the relation of displacement is concerned, the only property of the entity is that it occupies the terminal of the relation. The simple existence operators J, J' give a combined existence operator JJ', which will also be the existence operator of the displacement; for the relation will not exist unless both the relata exist.

We have now to form a continuum of displacements. This is constructed by forming a Group of operators, which includes all operators of the form JJ'. It is not necessary that the Group should consist of JJ' operators alone; but the JJ' operators will be distinguished as the pure operators of the Group.† It is simplest to consider the matrix representation of the

† In a Group the transformations consist of the same operators as the field. It would be contrary to the ordinary conceptions of transformation theory to limit the transformations to pure operators. Consequently the field must be extended to include impure operators.

operators. Since J has two eigenvalues, it can be represented as a twofold matrix. Then JJ' is represented by a fourfold matrix, or equivalently by an E-number. The required Group is therefore that of the E-numbers.

This is the simplest solution; but we must also show that more complex solutions are excluded. For example, J might be represented by an idempotent fourfold matrix; but such a matrix has been shown to be of the form $\frac{1}{2}(1+i\zeta).\frac{1}{2}(1+i\theta)$, i.e. the product of two idempotent twofold matrices. Since J is the product of two simpler existence operators, the entity defined by it can be analysed into two parts either of which can exist without the other. This case is excluded because it is implied in the conception of displacement that the entities related by it are like points "having no parts". To represent J by a matrix of order higher than the minimum order 2 implies that the entity defined by it is something more than an "occupied point".

It being established that displacement is represented by an E-number, we can proceed to develop the theory of space as in Chapters IV and VI. This gives the result that space-time is a curved four-dimensional continuum, and that its signature is $3+1$.

16·9. Reaction on Macroscopic Theory.

For the most part there is no occasion to modify the existing macroscopic relativity theory,† which I believe to be correct so far as it goes (except for the important amendment explained in § 8·8). The following are the chief points that claim attention when it is reviewed in the light of the present results:

(1) There would seem to be a doubt whether Einstein's theory holds rigorously for a rotating macroscopic body, e.g. a star.‡ The whole subject is very obscure, but I think the question can be put in the following form. Doubtless the particles in a star have attained a statistically steady distribution of spin, such that as many spin in one direction as in the opposite. But is this balance of spin relative to (*a*) axes rotating with the material, or (*b*) Galilean axes? Since the balance is brought about by interactions between the particles, either directly or with radiation as intermediary, I should conjecture that the answer is (*a*). If so we have, relative to the Galilean axes, a distribution of sources (amounting in the aggregate to a macroscopically appreciable source) for which Einstein's theory provides no notation. If the answer is (*b*), the difficulty does not arise.

† For definiteness, this is understood to be the theory set forth in my *Mathematical Theory of Relativity*, except that important advances have since been made in the cosmical problem. Chapter V should therefore be supplemented by R. C. Tolman's *Relativity, Thermodynamics and Cosmology*, pp. 331–488.

‡ Doubt of a similar kind has been raised by Sir J. Larmor (*Nature*, **137**, 271 (1936)).

Assuming (*a*), the case appears to be one in which it is necessary to use the Riemann-Christoffel matrix (§ 11·4). The effect, if any, can scarcely be large enough to be of macroscopic importance. Its interest lies in its bearing on the conservation of energy, since it suggests the existence of a new type of field which, if it is not taken into account, would render even the macroscopic conservation of energy imperfect.

(2) Since $g_{\mu\nu}$ and κ_σ denote measurable properties of macroscopic fields, they are expectation values determined by the wave functions of a great number of individual particles. We may indicate this average character explicitly by the notation $\overline{g}_{\mu\nu}$, $\overline{\kappa}_\sigma$. In general $\overline{g_{\mu\nu}\kappa_\sigma} \neq \overline{g}_{\mu\nu}\overline{\kappa}_\sigma$, so that in the generalised field theory it is represented by a separate symbol $K_{\mu\nu,\sigma}$.† The generalised theory reduces to Weyl's theory if $K_{\mu\nu,\sigma} = g_{\mu\nu}\kappa_\sigma$. Weyl's theory is therefore the approximation obtained by neglecting the difference between the average of a product and the product of averages. From a practical point of view the correction is trivial; but the question of practical application scarcely arose in these theories, whose aim was to gain insight into the foundations of world-structure. The point brought out in the general theory, which is not seen in Weyl's theory, is that the law of gravitation is not an additional limitation but is simply the gauging equation; this is one of the essential links in connecting the field theory with the microscopic theory developed in this book.

As it happens, we have employed Weyl's theory in a practical way on several occasions (§§ 12·7, 12·8). To study the electromagnetic field we create an artificial field of electromagnetic potential by a gauge transformation; just as in Einstein's theory we study the gravitational field by creating an artificial field of gravitational potential by a coordinate transformation.

16·95. Philosophical Outlook.

We conclude with a brief reference to the philosophical position towards which the present results trend. Unless the structure of the nucleus has a surprise in store for us, the conclusion seems plain—there is nothing in the whole system of laws of physics that cannot be deduced unambiguously from epistemological considerations. An intelligence, unacquainted with our universe, but acquainted with the system of thought by which the human mind interprets to itself the content of its sensory experience, should be able to attain all the knowledge of physics that we have attained by experiment. He would not deduce the particular events and objects of our experience, but he would deduce the generalisations we have based on them. For example, he would infer the existence and properties of radium, but not the dimensions of the earth.

The mind which tried to apprehend simultaneously the complexity of the

† *Mathematical Theory of Relativity*, § 93.

universe would be overwhelmed. Experience must be dealt with in bits; then a system must be devised for re-connecting the bits; and so on. One outcome of this treatment is that the universe is passed through a sieve with 3.10^{79} holes to render it more comprehensible. In the end what we comprehend about the universe is precisely that which we put into the universe to make it comprehensible.

So far as I can trace, the earliest sign of uneasiness among physicists about this procedure was shown in connection with the analysis of white light by a grating. The analysis of white light by a prism had been looked on as a discovery of its composite nature. When the same thing was done with a grating, it was seen that we were openly manufacturing the periodicities we thought we had discovered in it. As to the composite nature of white light, all that can be said is that if a mathematician chooses to analyse a function into Fourier components he is at liberty to do so.

The theory of relativity drew much fuller attention to the subjective aspect of many of the laws of physics. The Lorentz transformation comprises a number of laws which seem to describe properties of the natural objects which surround us (e.g. change of mass with velocity, FitzGerald contraction). But we put these properties into the objects because it is our habit to refer everything to a reference frame in which space and time are separated, although there is no such separation to be found in nature. Passing to a more advanced illustration of subjective influence, we have seen (§ 13·4) that the principle of least action arises because (as part of our system of comprehending experience in bits) we feel the need to *localise* the various measures which we employ. Localisation is an artificial concept in an interrelated universe; and indeed in elementary wave mechanics the conception of energy and momentum is primarily introduced as an attribute of infinite plane waves, spread over the whole universe.

In microscopic physics the question of how much we discover and how much we manufacture becomes still more acute. We cannot observe a microscopic particle without grossly interfering with it. It is often said that the particles are put into particular states by the type of experiment we perform on them. That is scarcely a fair view of the nature of the interference by the observer. Ideally he might wait until the conditions of the experiment reproduced themselves naturally. His interference is selective rather than active. But so far as physical theory is concerned, it makes little difference whether the observer selects the state, or puts the system into the state, which he "discovers". Thus before enumerating the characteristics of an elementary particle, we have to indicate the type of observation by which its existence is supposed to have been recognised. Was it the momentum or the position that was noticed? Or was it inferred as one of n particles whose total momentum and mean position were found?

Thus at various stages in this book we have had to distinguish specified, unspecified, and macroscopically specified particles, internal and external particles, neutral and vector particles and so on. Their diverse properties are the result of the different varieties of observational interference (active or selective) which precede our recognition of them. At the beginning we treated abstract elementary particles supposed to have been the subject of absolute measurements of position and momentum. In a sense these particles were all protons or electrons. But only when we recognised that actual measurements of momentum and position are relative, did we reach particles having the characteristic mass and charge of the proton and electron.

Fifteen years ago I was responsible for an oft-quoted remark,† "It is one thing for the human mind to extract from the phenomena of nature the laws which it has itself put into them; it may be a far harder thing to extract laws over which it has had no control. It is even possible that laws which have not their origin in the mind may be irrational, and we can never succeed in formulating them." This seems to be coming true, though not in the way that then suggested itself. Laws of atomicity have since been discovered, and have turned out to be rational and comprehensible to the mind; but it turns out also that they have been imposed by the mind in the same way as the other rational laws. But a new situation has arisen, because we now recognise that the totality of (mind-made) law does not impose determinism. Room is left within the scheme of physical law for undetermined behaviour. Behaviour whose laws are irrational was perhaps as near to the conception of undetermined behaviour as the thought of the time could reach.

The physicist might be likened to a scientific Procrustes, whose anthropological studies of the stature of travellers reveal the dimensions of the bed in which he has compelled them to sleep. Yet I do not think that we take unwarrantable liberties with the universe in our Procrustean treatment of it. If experience is a subject-object relation, the subject is entitled to—nay, he cannot divest himself of—his half-share. It can scarcely be a coincidence that Heisenberg's uncertainty principle has defined the half-way line with mathematical exactitude, distributing a coordinate to one side and a momentum to the other side with perfect impartiality. And so we may look forward with undiminished enthusiasm to learning in the coming years what lies hidden in the atomic nucleus—even though we suspect that it is hidden there by ourselves.

† *Space Time and Gravitation*, p. 200.

INDEX

A priori probability, 107, 198
Absolute and relative displacement, 109, 183
Action invariant, 136, 243
Action principle, 238, 240
Added system, 262, 266, 268
Addition, non-commutative, 154; generalised, 309
Additive characteristics, 154, 309, 315
Affine geometry, 183, 323
Algebraic coordinate in phase space, 106, 216, 220
Algebraic frame, 48
Algebraic function, 125; wave function, 120
Algebraic number, 20
Algebraic wave tensor, 89, 248
Angular momentum, 122; dynamical, 123; quantisation of, 147; macroscopic, 248, 326. See also *Spin*
Antedating, 111, 224
Anticommuting symbols, 21, 51, 312
Antiperpendicular rotation, 51, 94
Antisymmetrical matrices, 40, 96; double matrices, 159, 166
Antisymmetrical tensors, 54, 57, 58, 181
Antisymmetrical wave functions, 167, 288, 290
Antitetrad, 23, 69, 128
Antithetic, 75
Antitriad, 23, 69, 128
Associated strain vector and space vector, 103
Asterisk notation, 16
Augmented phase space, 217, 219
Average particles, 234

Basis of statistics, 198, 216, 281, 286
"Blank sheet", 32, 40, 56, 107, 109, 287
Bond, W. N., 272, 273, 303
Born, M., 151
Born-Infield theory, 136
Box problem, 235

Cardinal operator, 312
CD frame, 161, 167, 203
Chain multiplication, 16, 155
Chandrasekhar, S., 235
Characteristic equation, 44
Charge (electric), sign of, 49, 80, 88, 168; association with mass, 223; origin of, 283; value of e, 304
Charge-current vector, 90. See *Stream vector*
Clifford's numbers, 22
Closure of space, 78, 83, 97
Cogredient, 14
Collision interchange, 245
Combined systems, 110, 158, 215
Commutation rules, Jordan-Wigner, 312; Einstein-Bose, 314
Commuting operators (W, U_1, U_2, U_3), 128, 141; for hydrogen, 144
Compact E-numbers, 72, 101
Comparison fluid, 180; partial, 196; neutral, 206; indistinguishable, 294
Complementary stream vector, 311
Complete energy tensor, 163
Complete momentum vector, 65
Complete sets, 22, 35, 47, 156; transformation of, 27, 160
Complete space vector, 54; constitution of, 57
Complete stream vector, 65, 73
Complex conjugate wave functions, 131
Component of an E-number, 23; formula for, 37; in double frames, 157
Composite individuality, 234, 285, 307, 315
Configuration, 95
Conjugate triads, 23, 42, 47, 160
Conservation of probability, 115, 120; of energy and momentum, 229, 327
Constant of gravitation, 273
Constants of nature, 3; revision of, 304
Continuous and discrete wave functions, 213, 226
Contracting universe, 279
Contraction of matrix (spur), 36; of double matrix, 156; of wave tensor, 206; of volume element, 213
Contragredient, 14
Contravariant, 15
Cosmical constant (λ), 3, 188, 194, 279; indeterminacy of, 229, 238, 252
Cosmical Riemann-Christoffel tensor, 202, 210
Cosmical system (system A), 266
Coulomb energy, 281, 297, 301
Coupling of spin, 228
Covariant derivative, 121; with respect to χ, 289, 300
Covariant vector, 15; wave tensor, 19, 94, 109
Crossed frames, 161; significance of, 162
Crystal grating, 306
Curvature of space-time, 5, 55, 82, 126, 203, 256

Curvature tensor, gauge-invariant, 251
Cylindrical curvature, 220

D-symbols, 125, 317
Darwin, C. G., 1
de Sitter space-time, 5, 182, 192, 202, 207, 211, 264
Debye-Hückel effect, 139
Deflection value of e/m, 304
Degeneracy formula, 234, 253
Degeneracy of systems and eigenfunctions, 5, 247, 292, 321
Degenerate *E*-number, 23
Description, systems of, 13, 14, 117
Determinant of an *E*-number, 42, 107
Differential wave equation, 119; strain vector form of, 130
Diffraction of electrons, 305
Dirac, P. A. M., 1, 6, 15, 36, 62, 90, 115
Dirac's theory of the positron, 265
Dirac's wave equation, 63, 87, 106; differential equation, 119, 130, 144, 151
Discrete and continuous wave functions, 213, 226
Displacement vector, 15
Displacements, existence operators of, 325
Divergence operators, 118
Double frame, 155; wave vector, 155; phase space, 166; existence operator, 322, 325
Double tensors, importance of, 169, 181; reduction to simple tensors, 207, 212
Double-valuedness of wave function, 60, 151
du Val, P., 36
Dual of a tensor, 162
Dual R.C. tensor, 192
Dynamical coordinates, 116, 123, 127, 140
Dynamical equations, 126, 128, 144, 213; for strain vector, 136

E-numbers, 21; matrix representation (given in full), 42
E-symbols, 22; line as independence of, 24; representation by matrices, 36, 39
Eigenfunctions, enumeration of independent, 318; double and quadruple, 321
Eigensymbols and eigenvalues, 44
Einstein, A., 1
Einstein-Bose particles, 312
Einstein space-time, 5, 195, 211, 264
Einstein universe, 192, 256, 258, 278; metastability of, 277
Electric charge. See *Charge*
Electric moment, 89
Electrical energy tensor, 189, 194
Electrical matrices, 84; rotations, 84
Electromagnetic potential (κ_a), 119; gauge transformation of, 134; origin of, 137; macroscopic character of, 139
Electron-point, 140
Electron pressure, minimum, 234, 258
Elementary particle, 74
Energy, conservation of, 229, 327; furnished by specification, 197; threshold, 262
Energy invariants, 221, 260, 267
Energy levels of hydrogen, 147, 303; negative levels, 265
Energy operator ($\mathbf{T}_{\alpha\beta}$), 230, 250
Energy tensor, 91, 162, 169, 188, 229, 315
Environment, 32; standard, 129, 143, 150, 180
Epistemology, 5, 328
Equipartition of energy, 221, 254, 268
Equivalence of frames, 29, 32; of points in space-time, 56, 77; of volumes, 99
Exact energy tensor, 164
Exclusion principle, 231, 233, 253; limited to steady systems, 246, 254; in secondary quantum theory, 315
Existence, degree of, 309, 320
Existence operators, 308; double, 317, 322; geometrical representation, 320
Expanding universe, 225, 273, 279
Expectation value, 38, 149; of spin momentum, 122; of energy operator, 231, 257
Exponential, non-algebraic, 50
Extended wave function, 284
External and internal states, 92

Factorisation of wave tensors, 66; conditions for, 69, 70; in secondary quantum theory, 310, 314
Factors, numerical, $(\frac{1}{1}\,\frac{3}{3}\,\frac{7}{6})$, 10, 302, 304, 307; $(\frac{4}{3})$, 227, 258, 276; $(\frac{5}{3})$, 259, 276, 278; (2), 143, 152, 261, 274, 277, 318, 319
Factors, symbolic, 37
Fermi-Dirac statistics, 282; law of antisymmetry, 290; particles, 312
Field theories. See under *Generalised*, *Unified* and *Weyl's theory*
Final wave vectors, 16, 18
Fine structure constant, 3, 153, 295, 298
Five-dimensional relativity, 190
Five-dimensional space vectors, 53
Flat space-time, 5
Four dimensions, space vectors in, 55; reason for, 3, 325
Four-point elements of structure, 170, 323
Four-point matrices, 34, 38, 96
Fowler, R. H., 235
Frame, geometrical, 179

Frame, symbolic, 29; right- and left-handed, 48; neutral and macroscopic, 84; double, 155; crossed, 161; of identity, 291
Friedman-Lemaître theory, 279
Frobenius's theorem, 46

Gauge-invariant curvature tensor, 190, 251
Gauge transformation, 135, 223, 226, 295; non-integrable, 137; natural gauge, 136
Generalised field theory, 5, 85, 190, 295, 324, 327; amendment of, 136
Gravitation, constant of, 3, 273; law of, 136
Gravitational potential energy, 264
Ground state of the universe, 256, 259, 279
Group property, 13

Hamiltonian, 63, 129; reality conditions for, 87; strain vector form, 106; for internal coordinates, 175; for standing waves, 237
Heisenberg, W., 1. See *Uncertainty principle*
Helium, packing ratio in, 10
Hermitic conditions, 75, 131, 152
Homothetic, 75
"Hybrid" particles, 201
Hydrogen atom, wave equation of, 145; in practical units, 153; interaction in, 291; evaluation of Coulomb term, 298
Hyperbolic rotation, 76, 97, 217

Idempotent symbols, 70; physical importance of, 91
Identity, the fundamental, 67; generalisation of, 187
Identity (physical), relativity of, 290; composite, 234, 285, 307, 315
Imaginary gauge-transformation, 135
Imaginary matrices, 40
Index of wave tensor, 131, 152, 274, 304
Indistinguishable particles, 282; dynamics of, 285; comparison fluid, 293; protons indistinguishable from electrons, 287
Inhibited rotation, 110
Initial probability distribution, 107, 198, 216, 281, 286
Initial wave vectors, 16, 18
Inner product, 15; notation for, 17
Integrated energy tensor, 253
Interaction, 281; of system with environment, 112, 116, 129; at boundary, 249; in hydrogen atom, 291; in general systems, 154, 298
Interchange, continuous, 283
Interchange coordinate. See *Permutation coordinate*
Interchange energy, 245, 283, 293
Interchange operator, 158; factors of, 159; for double frames, 160; associated with Coulomb energy, 301
Interlocked transformations, 54, 109
Internal and external states, 92
Interval, 75; measurement of, 81, 242
Invariant time (s), 116, 129; represented by algebraic coordinate, 106, 216
Inverted cross, 161, 164
Irrational laws, 329
Irregular metric, 211
Irreversibility of time, 117, 225

Jordan, P., 151
Jordan-Wigner wave functions, 169, 311

K state of a system, 259, 269
Kinematical energy tensor, 189
Kinematical matrices, 84; rotations, 84
Kinetic energy, distribution law, 246
Kummer collineation group, 36

Left-handed frames, 48, 81, 85, 172
Lemaître, G., 47, 279, 319
Light-time, 111, 129, 218
Limit state, 259
Linear transformations, 13
Linkage of system to surroundings, 32, 111, 117, 129, 292
Linked rotations, 182, 187; translations, 189; displacement in time, 187, 192, 201
Littlewood, D. E., 71
Local orthogonal coordinates, 98
Local system (system B), 266
Localisation of characteristics, 240
Lorentz transformations, 76, 116, 249; inapplicable to internal state, 92, 96; to Einstein world, 211; to exclusion principle, 231; to non-transferable energy, 254

MacDuffee, C. C., 7
Macroscopic and microscopic theory, connection of, 4, 169, 201, 208, 256; in secondary quantum theory, 315
Macroscopic relativity theory, emendations required, 136, 279, 326; new light on, 190, 327
Macroscopic set, 83
Magnetic field, energy due to, 227; strong and weak fields, 228
Magnetic moment, 88, 89
Mass, origin of, 201, 262, 268, 276
Mass-constant, 230
Mass-ratio of proton and electron, 3, 219, 304

Matrices, 16; non-commutative property, 17; components, 36; real and imaginary, 40; Pauli, 47; kinematical and electrical, 84; space-like and time-like, 96; double, 155
Matrix representation, advantage of, 39, 96
Matrix representing general E-number (given in full), 42
Measurables, 170, 324
Measurement in strong fields, 81, 243
Metastable states, 150
Metrical tensor ($g_{\mu\nu}$), measurement of, 81; identified with kinematical self energy tensor of comparison fluid, 196, 209
Microscopic and macroscopic theory, connection of, 4, 169, 201, 208, 256; in secondary quantum theory, 315
Microscopic vorticity, 189, 326
Milner, S. R., 48
Minimum equation, 44
Minor complete set, 47
Minus particles, 88, 90, 264, 319
Mixed tensor, 14, 29, 54, 94
Modifying factor, 107, 281
Momentum and position, 162. See also *Uncertainty principle*
Momentum operator, 120, 131; in practical units, 152; relation to energy operator, 230
Momentum vector, 63; complete, 65; non-algebraic, 120
Monothetic, 75
Multiplication, inner, outer and chain, 15, 16; non-commutative, 17; of probabilities, 154, 213, 309
Multiplicity of frames, 204, 317
Mutual energy invariant, 221, 260; pressure invariant, 221, 231
Mutual energy tensor, 165, 205, 315

Natural coordinates, 98
Natural gauge, 136
Nebulae, speed of recession, 279
Negative and positive charges, 81, 88, 223, 301; balancing of, 278
Negative energy levels, 265
Negatrons, 90, 265
Neumann, J. v., 2, 36
Neutral particle, 89; mass of, 219, 267; formula for mass (m_0), 271, 275
Neutral space-time, 80; sets, 83; comparison fluid, 206
Neutron, 7
Newman, M. H. A., 42
Nonads, 42
Non-algebraic phases, 97; wave functions, 120
Non-commutative multiplication, 17; exponentials, $5v$, 53; addition, 154; wave functions, 311
Non-integrable gauge transformation, 137
Normalised E-numbers, 70; strain vectors, 106
Nuclei, atomic, 7; disruption by protons, 246
Number of dimensions of space-time, 3, 325
Number of particles in the universe, 272, 316, 325; relation to number of eigenfunctions, 319
Numerical constants of nature, 3

Object system, 180
Observables, 170, 324
Observer, 79; his selective interference, 328
Outer product, 15; notation for, 17

Packing ratio in helium, 10
Parallel displacement, 122; generalised, 126; in comparison fluid, 182; in χ, 288
Parallel strain, 134
Parameters of wave functions, 60; promotion to coordinate rank, 123, 289
Partial comparison fluid, 196, 198, 207, 210
Partial existence, 309, 320
Partially occupied states, 311, 317
Particles, 8; elementary, 74; neutral, 89; scalar, 231; number in the universe, 272, 316, 325
Particles, external and internal, 174, 221; of composite identity, 234, 285, 307, 315; indistinguishable, 282, 285; excluding (Fermi-Dirac) and non-excluding (Einstein-Bose), 312
Partition, scheme of, 183
Pauli matrices, 47, 160, 248, 278, 310
Pauli space, duplexity of, 321
Pentad, 23, 35; real and imaginary members, 40
Pentadic part, 68
Permanent, 59
Permutation coordinate, 153, 228, 283, 284; relativity of, 291; recoil, 293
Perturbation, 116, 141, 239, 249, 292
Phase, 97
Phase space, 95; singular, 112; double, 165; in *CD* frame, 166
Philosophical outlook, 327
Photons, 263, 315
Physical reality, invariance of, 76
Position and momentum, 162. See also *Uncertainty principle*
Position vector, 73, 162; tensor, 164
Positive and negative charges, 81, 88, 223, 301; balancing of, 278

Positively saturated space, 81
Positron, 90, 264
Potential. See *Electromagnetic potential*
Potential energy, 245, 264
Pressure in degenerate gas, 235, 236; relation to energy-density, 237, 241; change of zero reckoning, 258, 269
Pressure invariant, 221, 231
Probability, 107; distributions of ("states") 32, 115; conditions of conservation, 119, 120; concerned in dynamical equations, 142; multiplication of, 154, 213; of identification, 284, 286; form of partial existence, 309. See also *Initial probability distribution*
Procrustean methods, 329
Progressive waves, 247
Projective relativity, 56
Proper time. See *Invariant time*
Pseudo-reciprocal, 26, 63, 112
Pure wave tensors, 65, 68; standard forms, 69; idempotent condition, 70, 73; physical importance of, 91
Purity lost by contraction, 208

Quadratic equation for mass, 219, 304
Quadruple wave functions, 170, 297, 323
Quantum numbers, n', 147; u, 147; μ, 149
Quantum theory. See *Microscopic theory* and *Wave mechanics*
Quarterspur, 23, 37
Quotient of E-numbers, 25

Radiation, 7; fields of, 143, 262
Real and imaginary matrices, 40
Reality conditions, 86; in phase space, 97; for gauge transformation, 135
Recession of the nebulae, 279
Reciprocal, 26
Recoil of comparison fluid, 183; in time, 186, 191; in χ, 293; in r, 296
Reduced mass, 174
Reflection of a complete set, 28
Relata and relations, 170, 181, 323
Relative and absolute displacement, 109, 183
Relative coordinates, 170, 174; energy tensor, 197, 205; time, 217, 295; identity, 290; existence operator, 313
Relativistic degeneracy, 235, 246, 253
Relativity rotation (or transformation), 30, 50, 155, 248; of identity, 283, 285
Rest energy (or mass), origin of, 262, 268, 276
Riemann-Christoffel matrix, 187
Riemann-Christoffel tensor, 181; relation to energy tensor, 188; dual of, 192; cosmical, 202
Right- and left-handed frames, 48, 81, 85, 172
Rotation, 30; antiperpendicular, 51, 94; circular and hyperbolic, 76, 97, 217. See also *Relativity rotation*

Saturated space, 81, 164
Scalar particles, 231; mass-constant of, 236; system of N', 259
Scalar product, 15; notation for, 17
Scalar wave functions, 229; relation to vector wave functions, 247
Schrödinger's wave mechanics, 75, 89, 143, 229, 232
SD-numbers, 35
Second quantisation, 308
Secondary quantum theory, 310, 315
Self-consistent field, 260, 312
Self-energy tensor, 162, 165, 205, 209, 315; complete, 163
Self-pressure and energy invariants, 221, 260
Signature of space-time, 3, 326
Simultaneity in internal space, 92, 110; not applicable to imaginary time, 295
Single-valued wave functions, 150, 287
Singular matrices and symbols, 18, 26; conditions for, 43, 44
Singular phase space, 112; transformations, 173, 308
Sitter, W. de. See *de Sitter*
Six-dimensional group of rotations, 85
Six-vector, 57
Sommerfeld's formula for energy-levels, 147, 302
Sound waves, 245
Space-like matrices, 96; in double frame, 165
Space tensors, 15; vectors, 53; complete, 54; association with strain vectors, 104, 110
Space-time, four dimensions of, 3, 325; geneses of macroscopic, 80
Specification of particles, 196; energy furnished by, 197, 267; macroscopic, 201
Spectral set, 72; analysis of algebraic numbers, 89, 249
Spherical curvature (Gaussian), 78, 220
Spherical space, 5, 55. See *de Sitter* and *Einstein space-time*
Spin, 73, 88; momentum, 122; coordinates, 221, 283; duplexity, 321
Spur, 36
Standard environment, 129, 143, 150, 180
Standing waves, 234, 244; hamiltonian of, 237
States, 32, 115; degeneracy of, 124, 247, 289, 292
Statistical equilibrium, 141, 239

Steady state, 140; generalised, 143
Stereographic coordinates, 101; application of, 218, 219
Stereographic projection, 257
Stern-Gerlach effect, 226, 278; in weak fields, 228
Straight cross, 161, 164
Strain vector, 94; association with space vector, 104, 110
Strain vector form of wave equation, 106, 130
Stream vector, 65; of elementary particle, 88; connection with momentum, 122; algebraic, 248; in secondary quantum theory, 310
Strong action principles, 242
Strong and weak magnetic fields, 228
Suffix coordinates, 283
Suffixes, omission of, 16
Summation convention, limited use, 22
Symbolic calculus, 20, 125
Symbolic factors, 37
Symbolic frame, 29
Systems A, B (cosmical and local), 266; A', B', 270

Temple, G., 27, 73, 145, 151
Tensor, 13
Tensor-density, 58
Tetrode, H., 2
Threshold energy, 262
Time, distinctive character of, 75; in internal states, 92; irreversibility of, 117, 225; linkage of displacement in, 187, 192; treated differently from spatial displacement, 201. See also *Invariant time*
Time, imaginary relative, 295
Time-like matrices, 96; in double frame, 165
Tolman, R. C., 256, 273, 326
Transformations, etymological, 136
Transformations, linear, 13; of wave tensors, 17, 18; of complete orthogonal sets, 27; unitary, 43, 98; of strain vectors, 94; general symbolic, 125; of gauge, 134; non-integrable, 137; of double wave vectors, 155; of double frames, 162; associated with interchange, 283, 287; of existence conditions, 308. See also *Lorentz transformations*
Transitions, 185, 292
Translations, 77; linkage of, 189
Transpose, 19
Triads, 23; conjugate, 23, 42, 47, 160
Two particles, representations of, 167

Uncertain energy tensor, 164, 203, 210
Uncertainty principle, 74, 162, 169, 179, 191, 275, 286, 329
Unified theories, 4, 243
Unit matrix, 18
Unitary transformation, 43, 98
Universe, 7; Einstein and de Sitter forms, 211; analysis of Einstein universe, 256; expanding, 273, 279; number of particles in, 272, 316, 325
Unspecified particles, 196

Vector, 16; basic, 15, 54; space, 53, 54; strain, 94; double, 155
Vector density, 58; three-dimensional, 59, 105; discriminated by uncertainty principle, 74; by reality conditions, 86, 87
Vector wave function, 60; relation to scalar wave function, 248
Velocity of indistinguishable particles, 285
Volume element, 58, 99; contraction of, 213
Volume of phase space, 99; of spherical space, 257

Wave equation, 1, 62, 106; differential equation, 119, 130; for hydrogen, 144, 153, 298
Wave functions, 60, 66; replacement of double by single, 212; scalar, 229; n-tuple, 244; Jordan-Wigner, 310
Wave mechanics, 1; method, not theory, 7, 66, 256; importance of conception of states, 32, 115; principles of approximation, 112, 129, 249; a statistical theory, 239; modified outlook of secondary theory, 309, 315
Wave packets, 225, 263, 286; pockets, 263
Wave tensor calculus, origin of, 2, 15
Waves equivalent to curvature, 256
Weak action principles, 242
Weak and strong magnetic fields, 228
Weyl's theory, 1, 5, 85, 135, 136, 252, 295, 327. For the author's generalisation see *Generalised field theory*
White dwarf stars, 235, 255

Zanstra, H., 10
Zariski, O., 36
Zero reckoning of energy and pressure, 188, 195, 197, 229, 238, 258, 266, 313

CAMBRIDGE: PRINTED BY WALTER LEWIS, M.A., AT THE UNIVERSITY PRESS