DAMAGE BOOK

UNIVERSAL
LIBRARY

RELATIVITY
FOR PHYSICS STUDENTS

BY

G. B. JEFFERY, M.A., D.Sc.
FELLOW OF UNIVERSITY COLLEGE, LONDON, AND PROFESSOR OF
MATHEMATICS AT KING'S COLLEGE, LONDON

WITH FOUR DIAGRAMS

METHUEN & CO. LTD.
36 ESSEX STREET W.C
LONDON

First Published in 1924

PRINTED IN GREAT BRITAIN

CONTENTS

RELATIVITY FOR PHYSICS STUDENTS

EINSTEIN'S THEORY OF RELATIVITY*

AS I conceive the office of a professor, it is that he should stand before his students as the living representative of those great men who in the past have laboured in that branch of human knowledge which he has made his own; that by means of a reverent yet unflinching criticism he should strive to reveal the workings of these master minds, to the end that he may impart, not merely knowledge, but that more precious gift—the art of acquiring knowledge, the art of discovery. If we

* An Inaugural Lecture, delivered at King's College, London, on 9*th October*, 1922.

approach our task in this spirit, we shall find the key to the solution of much that is difficult and perplexing in our present knowledge, and the inspiration which will lead us on to further discoveries.

It seems natural, therefore, that I should seek to illustrate this theme by means of the subject which throughout my mathematical career has inspired me more than any other branch of mathematics or physics into which my work has led me, and the subject which, as far as one may venture to prophesy as to the future course of scientific thought, seems marked out for great advances in the immediate future.

Einstein's theory of relativity has proved full of difficulty to layman and expert alike. It seems to invite us to cut ourselves loose from all that has gone before, to scrap all our old ideas and to start afresh with new. It is true that the theory does profoundly modify our fundamental ideas of space, time, and motion, but a deeper study reveals the fact that it is nevertheless the natural and almost inevitable sequel to the work of the great masters of the past, and more particularly to the work of Isaac Newto himself; how natural and inevitable it wi be the main purpose of this lecture to show.

The story of modern mechanics begins in the sixteenth century. Tycho Brahe, with no telescope, and the most primitive instruments in place of the equipment of the modern astronomical observatory, sustained through years of labour by a most extraordinary patience, observed night after night the positions of the planets among the surrounding stars. Tycho was one who sowed but did not reap. Those who hav any experience of observational astrono find it difficult to imagine a duller task Tycho's—the accumulation of volu figures whose meaning it was not him to read—the construction of Almanac without its beautiful order. Nevertheless his work the necessary foundation fc followed.

The task was taken up by Tycho's pupil and assistant, John Kepler. He succeeded in clothing his master's data with the form of three simple descriptive laws. His was a great achievement. All Tycho's volumes of figures, all those strange motions of the bodies which men have most appropriately called wanderers, were summed up in three simple statements. Kepler had no theory; he made no attempt to explain the motions he studied. The Archangels who kept the celestial spheres in motion were dismissed, but no subtle scientific hypothesis was imported to perform their office. In effect he said: Viewed from the earth, the motions the planets are very complicated. Now wander forwards; now they retrace teps; now they move in loops; and ong sweeping curves. But viewed un, these motions are very simple, ch planet describes a perfectly

alileo had brought the same on the problems of terrestrial dynamics. He investigated t... which govern the motions of falling bodies. It had been taught that every body had its "proper place." The proper place of heavy bodies was low down, and the proper place of light bodies was high up. A body tended to move to its proper place; the heavier a body, the more quickly it fell, since it was presumably at a greater distance from its proper place. It is a striking commentary on medieval thought, that it seems to have occurred to nobody before the time of Galileo to test this conclusion by means of a simple experiment. Galileo made such an experiment at the leaning tower of Pisa, and found that all bodies, heavy or light, f... towards the ground in precisely the s... way. By careful laboratory experi... he ascertained the law of this fa... body falls towards the ground wi... which increases in proportion ... so that its speed is increas... 32 feet per second in every ... fall.

, Newton found ready to hand two sets of descriptive laws: Kepler's laws, which embraced the motions of the planets; and Galileo's law, which covered a very important case of the motion of terrestrial bodies. His first step was to throw the laws of Kepler into a different form. No doubt he took the hint from Galileo's law of falling bodies, and he investigated the motion of a planet, moving in accordance with Kepler's laws, from the point of view of the change of its velocity, or, as we should say, its acceleration. He found that Kepler's laws are equivalent to the statement that the acceleration of a planet is always rectly towards the sun, and that this leration depends in no way on the but only on its distance from the inishing with increasing distance in with the law of the inverse

served the similarity between the planets round the sun of bodies falling towards the earth. They too fall with an acceleration which in no way depends on the falling body. Is this gravitation subject to the same laws as the gravitation of the sun which keeps the planets in their orbits? Does it also diminish as the inverse square of the distance? It is difficult to answer these questions in the narrow range of height we can employ at the earth's surface, but Newton took the heavens for his laboratory. The moon, though somewhat disturbed by the sun, moves round the earth approximately in accordance with Kepler's laws, and has an acceleration towards the earth. Is this acceleration just what 32 feet per second per second would become if it diminished in accordance with the inverse square law up to the moon's distance? Newton worked the sum and found that it was so.

Thus the inward nature of gravitation was laid bare. There is a gravitation of the sun, in virtue of which *any* planet, comet, or meteorite which may happen to find itself in a given position experiences an acceleration which depends only upon that position. There is a gravitation of the earth, in virtue of which the moon, or any unsupported body near the earth, experiences an acceleration which again depends only upon the position of the accelerated body.

Thus far Newton was on very safe ground, for he was merely expressing the results of observation in a concise and compact form. He then proceeded to frame a theory which should account for the observed facts. Here we can trace the influence of Galileo very clearly. From his experiments on the motion of a body down an inclined plane, Galileo inferred that a body moving on a horizontal plane would continue to move with a constant velocity in a straight line. Earlier thinkers had felt the necessity of ascribing some cause to the motion of bodies ; if a body moves some agency must be at work to maintain its motion. The experiments of Galileo, and Newton's interpretation of Kepler's laws, conspired to promote the view that it was the *change* of motion, the acceleration of a body, for which a cause must be found, rather than the motion itself. Newton adopted the view that when the motion of a body changes it does so because the body is acted upon by a force, and that this force is measured by the product of the mass of the body and its acceleration. Gravitation is explained by the action of forces arising from, and directed towards, attracting bodies. This in the barest outline is the Newtonian system of mechanics as commonly understood. Before we proceed to criticise it, it may be well for a moment to pause to consider the achievement which stands to its credit. The motions of the planets are not, in fact, quite so simple as the laws of Kepler would indicate. Someone has said that if Kepler had possessed a modern telescope he would never have discovered his laws. Nevertheless, with a few small outstanding differences, the deviations from Kepler's laws are all explained when we take into account the gravitation of the planets upon each other. The history of dynamical astronomy has been very largely the verification, to an ever-increasing degree of refinement, of Newton's law of universal gravitation. Cavendish observed the workings of this same law in the attraction between quite small bodies in the laboratory. The laws of motion, originally deduced from the motions of the planets, are verified day by day in every engineering workshop.

It seems to me that the supposed conflict between Newton and Einstein rests very largely upon a failure to apprehend a distinction upon which Newton was always insisting, the distinction between what he called mathematical principles and philosophical principles. Mathematical principles were to Newton, not ultimate causes, but merely concise descriptions of the phenomena of Nature, which could be verified by observation and experiment. He distinguishes them very clearly from philosophical principles, whose function it is to explain and interpret phenomena. This distinction, maintained in actual scientific work, is one of the great debts which we owe to Newton. It defines at once the purpose and the limitation of Science. When Science shall have accomplished its purpose and described the whole material universe in the simplest way, it must leave us face to face with the philosophical problem of the mystery and meaning of the things which it has described. But Newton, like many of us, had within him something of the philosopher. He might jeer at the metaphysicians, but at times he could not help speculating, and rightly speculating, as to the meaning of those great descriptive laws which he found running throughout the whole fabric of Nature. He was, however, always careful to distinguish these speculations from the formulation of the mathematical principles which he regarded as the main part of his work, and we find them for the most part in the scholia in the

Principia, and in the queries in the Optics. These speculations have been the subject of controversy ever since, and it is towards them that the criticism of Relativity is, for the most part, directed.

In a scholium which follows the definitions in the Principia, Newton sets forth his views on time, space, and motion. He distinguishes between absolute time and relative time which is measured by some motion. He says :—

> "The natural days, which, commonly, for the purpose of the measurement of time, are held as equal, are in reality unequal. Astronomers correct this inequality, in order that they may measure by a truer time the celestial motions. It may be that there is no equable motion, by which time can accurately be measured. All motions can be accelerated or retarded. But the flow of absolute time cannot be changed. Duration, or the persistent existence of things, is always the

> same, whether motions be swift or slow or null."

In the same way he distinguishes between absolute space and relative space, and between absolute motion and relative motion. He says :—

> " We use in common affairs, instead of *absolute* places and motions, *relative* ones ; and this without any inconvenience. But in physical disquisitions, we should abstract from the senses. For it may be that there is no body really at rest, to which the places and motions of others can be referred."

Thus we need go no further than Newton himself, to find a clear statement of the problem to which the theory of relativity has attempted to supply an answer. Our experience is entirely of relative motions. We are at rest relatively to our immediate surroundings ; we are moving at a rate of 100,000 miles an hour relatively to the sun ; we are moving relatively to Sirius at such and such a speed; but how we are moving in an *absolute* sense, without reference to any other body, is a question which experimental science has often tried, but always failed, to answer. The statement that we are moving at a rate of 100,000 miles an hour is devoid of all physical meaning whatsoever, unless we state what we conceive to be at rest. This something, which for a particular purpose we assume to be at rest, we call our "frame of reference."

Now, if we consider Newton's work in its proper setting, there is no doubt at all as to what his frame of reference was. It was implicit in Tycho's data, and Tycho observed the motions of the planets relatively to the fixed stars. Newton's frame of reference was one in which the distant fixed stars are at rest. It seems likely that Newton, who boasted that he did not frame hypotheses, adopted the hypothesis of absolute space because in the fixed stars he found ready to hand a frame of reference which transcended the domestic motions of the solar system—the chief objects of his study. Was not Newton's absolute space after all just the physical space mapped out by the fixed stars, rather than the metaphysical concept we have usually taken it to be?

In the light of modern knowledge this frame of reference presents great difficulties. We can now, in many cases, measure the velocities of these stars relative to each other and to our sun. We find that they are not fixed, or at least, they are not all fixed, for they move relatively to each other with widely different velocities. The reason why, night after night, they seem to occupy the same positions in their constellations is the same as that which makes an express train seem to move so slowly when viewed from a long distance across country. It is not that their motions are slow—in many cases they are almost inconceivably great—but that the stars themselves are at such immense distances from us. Still more modern knowledge forbids us to attempt to surmount this difficulty by supposing that the motions of the stars are random, like the motions of the atoms of a gas, so that we could average them out, in order to arrive at our fixed frame of reference. If our stellar system has indeed grown out of a giant nebula, there may be an ordered system in the motions of the stars.

By the time that the discordant motions of the stars had been well established, a new hope had arisen. The undulatory theory of light seemed to call for some medium to transmit the light vibrations, and the idea of an ether pervading all space was developed. Clerk Maxwell showed the intimate relation between light and electromagnetism. Later on, the electron theory promised to explain the whole of physics in terms of electricity. Matter was simply an aggregation of electric charge, and electricity was a state or singularity of the ether. The ether had become fundamental in physics. Here it seemed that the solution of all our difficulties might lie. A body moves when it moves relatively to the ether; our frame of reference is to be fixed, not with respect to the so-called fixed stars, but with respect to the ether.

The result did not work out happily. If mechanics adopted the ether in order to simplify the problem of motion, never was foster-parent blessed with a more unruly child. If we observe a star, the ether is undisturbed by the earth's motion through it; if we fill our telescope with water, the water communicates part of its motion to the ether; if we make an interference experiment in the laboratory, we can only conclude that the earth carries the ether with it in its motion. Quite apart from the logical difficulty as to how the ether, the standard of absolute rest, can itself move at all, it moves or it does not move in a delicate accommodation to the particular experiment which we may happen to have in hand at the moment. In spite of the labours of some of the greatest English mathematicians of the latter half of the nineteenth century, the situation grew steadily worse.

In the meantime experimental physicists had concentrated on the problem of the determination of our motion relative to the ether. Many different experiments were proposed and carried out with all the skill and ingenuity of a great generation of experimenters. The result was always the same. No experiment succeeded in revealing our motion through the ether. The story is not unlike that of an earlier chapter in the history of science, which tells how for centuries men tried to construct a perpetual motion machine. They failed, and out of their failure modern physics has erected a great principle. They searched in vain, until they were led to deny the very possibility of the thing they sought. That denial has become the Second Law of Thermodynamics, one of the most powerful principles of modern physics. Relativity is the outcome of the application of the same method to our present difficulty. As the result of repeated failure, it asserts that no physical experiment can ever reveal our motion through the ether.

This was the culmination of a long sustained effort to bring the absolute space of Newton within reach of physical experiment, or perhaps we should say, rather, to restore to absolute space the physical reality which it lost on the discovery of the motions of the fixed stars. It is the starting point of the theory of relativity, that no method has yet been discovered by which this can be accomplished. If this position is accepted it constitutes a fatal criticism of Newton's laws of motion, at any rate in the form in which he stated them, for the very terms of those laws—motion, change of motion—have no meaning apart from some pre-determined standard of rest or frame of reference. It is obvious that the time had arrived when some fundamental reconstruction of the theory could no longer be delayed. Einstein did not bring forth his theory merely as an elaboration and refinement of physical law in order to bring theory into accord with a few isolated and newly-discovered facts; he brought it forth to meet the situation created by a complete theoretical breakdown of the older system.

If we seek a way out of the difficulty, the first suggestion which presents itself is that, since our experience is confined to relative motions, it ought to be possible to express the laws of motion in terms of relative motions alone, without any reference to absolute motions.

This in effect is what Einstein has done, though he approached the problem from a rather different point of view. If we take any frame of reference, we can obtain laws which will describe the course of natural phenomena. Since we have to recognize that the choice of a frame of reference is arbitrary, we shall expect these descriptive laws to be different if we choose another frame of reference. In other words, we shall expect to find that our descriptive laws are relative to the particular frame of reference which we have chosen. For example, if we take a frame of reference fixed with respect to the earth, we shall obtain the Ptolemaic system of astronomy with its epicycles, etc., whereas if we take a frame of reference fixed with respect to the sun, we shall obtain the very different descriptive laws of Kepler.

The question to which Einstein addresses himself is, whether the descriptive laws of physics can be framed in such a way that if they are true for one frame of reference they will also be true for any frame of reference whatever. This is essentially a mathematical question. If it is answered in the affirmative, the experimental question will arise as to whether these general laws are in fact true for one frame of reference. By the aid of the calculus of tensors, Einstein was able to give an answer to the mathematical question, and it appears that it is possible to frame laws which are absolute in the sense that, if they are true at all, they are true independently of the particular frame of reference which we may happen to choose. If these laws are verified by experiment, we shall have succeeded in dispensing with absolute space and with all the difficulties to which the introduction of this concept into our scientific work has given rise.

As so often happens in scientific research, Einstein's efforts to clarify our fundamental ideas of mechanics led to an important extension of knowledge. He was able for the first time to bring gravitation into relation with other physical phenomena. Let us return for a moment to the view of gravitation which we have already considered. In the Solar system, and in the fall of heavy bodies towards the earth, we observe the same essential feature, namely, that any body placed in a particular position experiences an acceleration which depends in no way upon itself, but only upon the position in which it is placed. In his determination to confine himself to the description of phenomena, Einstein accordingly regards gravitation as a property of space varying from place to place, leaving open for the time being the question as to whether this property can be expressed in terms of the influence of attracting bodies. In this sense Einstein's space, unlike that of Newton, is not homogeneous, but differs in its properties from place to place.

We can best explain Einstein's discovery by means of a simple, if somewhat fanciful, illustration. Imagine a lift working in a deep well, and let it be one of the kind which is operated, not from within the cage, but by a man at the bottom. Suppose that within the lift is the ghost of Galileo. He will be unconscious of the mechanism of modern lifts, but he might well return to his old task of the investigation of the laws of falling bodies. This he might do by allowing a marble to fall through a measured height to the floor of the lift and timing its fall. To avoid complications, we will allow him a stop-watch in place of his water-clock. The ghost sits there all day long, condemned to time the fall of this marble over and over again. So long as the lift remains stationary he will get the same answer every time. But suddenly he finds that the marble is falling more quickly, and he will say that gravity has increased. The man at the bottom knows better. He is sending the lift upwards with an accelerating speed. The floor of the lift is rising to meet the marble, and thus the latter accomplishes its measured journey more quickly. The man has only to make the lift go upwards or downwards with the right acceleration in order to make the ghost's gravity anything he pleases, downwards or upwards, or nil. If the man chooses to play tricks by sending the lift now up and now down, the poor ghost will find that gravity is fluctuating wildly, and will think that some kind of gravity storm is in progress. But the man on solid earth at the bottom knows that it is all due to the motion of the lift. If only the ghost would realize that he is being fooled, and that his frame of reference is being accelerated upwards and downwards, he would see that gravity has remained the same all the time. Einstein is inclined to make allowances for the ghost. He claims the liberty to take any frame of reference he pleases, and he is prepared to allow that the ghost was perfectly reasonable in taking himself and his immediate surroundings as a frame of reference. This fact, however, emerges, that if the same phenomenon is described from the point of view of different frames of reference, the gravitation inferred will, in general, be different. This is the essence of Einstein's equivalence hypothesis. He describes a physical phenomenon in the absence of gravitation by means of an accelerated frame of reference, and thus obtains a description of the same phenomenon in the presence of gravitation. It was by this method that he was able to establish the influence of gravitation on the propagation of light.

It is interesting to note how near Newton got to this idea. He lived much closer to the Copernican controversy than we do. Men had only just given up the belief of centuries that the stars revolved in their courses once a day. In his " System of the World " we find him facing the problem that the choice between the Ptolemaic and Copernican systems could not be settled by observation alone, but he points out that, whereas on the Copernican system gravitation can be expressed in terms of forces directed towards definite bodies which may be regarded as the sources of the gravitation, on the Ptolemaic theory the forces would be directed, not to the earth, but to points on the axis of the earth. He says :—

> " That forces should be directed to no body on which they physically depend, but to innumerable imaginary points on the axe of the earth, is an hypothesis too incongruous. 'Tis more incongruous still that those forces should increase exactly in proportion of the distance from this axe. For this is an indication of

an increase to immensity, or rather infinity; whereas the forces of natural things commonly decrease in receding from the fountain from which they flow."

Newton adopted the Copernican frame of reference, not on observational grounds, but because that frame of reference possessed the peculiar convenience that it enabled him to express gravitation in a simple way as arising from the influence of attracting bodies.

It should be pointed out that Einstein does not say that it is a matter of indifference as to which frame of reference we adopt. A man who attempted to conduct experiments in an unsprung cart, and who took the body of his cart for his frame of reference, would be asking for trouble, for he would have to deal with a hopelessly complicated gravitational field. The importance of the principle lies in this—that while one frame of reference may be more convenient than another for the discussion of some particular problem, all frames of reference are theoretically admissible.

Another consequence of the denial of absolute motion has been to destroy the independence of space and time. There has been so much misunderstanding on this point that it may be well to state exactly what relativity has to say on the matter. It may be stated very simply thus: At two distant points there is no definite unique instant of time at the second which may be regarded as simultaneous with a given instant at the first. For example, a new star bursts suddenly into view, and the astronomers tell us that, owing to its great distance and the time that it takes light to travel from it to us, the cataclysm which has made it visible must have occurred in the time of Newton. Such a statement would necessarily be approximate, for we have only the roughest notion of the distances of stars so remote as this one would have to be. But let us in imagination concede the astronomer all the accuracy of his wildest dreams. Could he even then assure us, for example, that the cataclysm occurred at the precise instant at which the famous apple struck the ground ? No, for perchance the solar system is moving in the direction of this star with a speed which we may very appropriately call *x*. If so, we are rushing to meet the light on its journey towards us, and we shall receive it sooner. How much sooner will depend upon *x*, and *x* has no meaning apart from a frame of reference. Adopt a frame of reference in which we are moving in the direction of the star, and the apple fell before the star burst into flame. Simultaneous and the words before and after, as applied to two instants of time at different points of space, have no precise scientific meaning apart from a specified frame of reference. Thus the time of one frame of reference depends upon the time and the space of another frame of reference. In the words of Minkowski : " Time of itself, and space of itself, fade into shadows, and only a kind of union of the two shall maintain an independent reality."

Thus the new theory has worked a fundamental change in the concepts of space and time. With Newton they were independent, homogeneous, absolute, and infinite; with Einstein they are but different aspects of the same continuum, space-time—heterogeneous, relative, and possibly finite.

It is often objected that relativity purports to disprove the existence of the ether, and that without the ether phenomena such as the propagation of light are inconceivable. It is not certain that relativity does do this. What has been shown is that the ether cannot be made to provide a standard of rest, and that the idea of motion of the ether is self-contradictory. This may mean no more than that the ether is a reality to which the idea of motion cannot be applied. It may be helpful to remember that precisely the same criticism was directed against Newton by the Cartesians. Because he refused to be drawn into discussions as to the plenum and its vortices, he was made to appear to say that the forces of gravitation were transmitted through emptiness from one heavenly body to another. Now it is quite clear that Newton's space was more than mere nothingness, in that it acted as the medium for the transmission of gravitational influences. Yet Newton was right in regarding the nature of this space, except in so far as it was susceptible to physical measurement, as a problem for philosophy rather than for science. The present position of the problem of space-time and the ether is, I think, very similar.

Time prevents us from referring to the practical achievements of the new theory or from exploring its possibilities in the regions in which the Newtonian mechanics have never yet shed light—the regions of atomic and sub-atomic structure. The formal beauty of the theory can only be exhibited by means of mathematical analysis.

I said at the beginning of the lecture that the record of the past would provide the key to the solution of much that is difficult and perplexing in our present knowledge, and I hope that, by attempting to put Einstein's work into its proper historical setting, I have perhaps made some aspects of the theory of relativity a little clearer. But I also suggested that the record of the past would point us on the way to further advance. The work of Newton was carried on by the great French school of the Revolution period. He laid down the principles, but it was Lagrange, Laplace, Poisson, and others who reduced them to a form in which they could readily be applied to the solution of physical problems. Again Einstein has given us the principles, but it is not always easy to see how to apply them to all those problems of modern physics which are so urgently with us to-day. That is the task which now lies before mathematics. Einstein has given us the "Principia," but "La Mécanique Analytique" has yet to be written.

I

THE ORIGINS OF THE THEORY

THE young scientist can suffer from no greater fault than a misunderstanding of scientific genius. We are too apt to think that advance in scientific knowledge is reserved for those who, by reason of some special gift, are able for the first time in human history to see something which others have been too blind to see. Now there is a measure of truth in this view, but it is important that we should see just what that measure is, more particularly when we come to the conclusion, as most of us must quite young in life, that we are very ordinary people with no very special gifts. When that moment comes, it will depend upon our conception of the way in which science advances whether we go forward and play our part, such as it may be, in the progress of knowledge, or whether in despair we leave the matter to those more fortunate ones who have been predestinated for the work.

The view of scientific advance which I wish to combat may perhaps be explained by an analogy. A region of country has been explored and mapped. The rivers and mountains, since they must have names, are called by the names of those who first discovered them. We have some knowledge of the geography of the land and perhaps we have reached the boundary. There we stand facing a mountain precipice, vainly seeking some way by which we may get a little further, and all the time hoping that some super-man may invent an aeroplane of thought which shall carry us over into the beyond.

The analogy is incomplete, for it leaves out of account the most important thing of all—the way in which what is now known was first discovered. The map may be sufficient for the engineer, but the discoverer must study much more than the map: he must be learned in the art and lore of exploration. Thus if we would become discoverers in physics, or even if we would in any real sense understand physics as it is to-day, the names of Galileo, Huygens, Newton, Lagrange, Fresnel, Stokes, Maxwell must be much more to us than convenient labels for certain laws and experiments.

For this reason we will approach our study of Relativity by showing how the problems which it attempts to answer have gradually arisen. There is another reason which prompts us to adopt this course. Relativity is commonly supposed to be a revolutionary theory. The theory has its roots in the very beginnings of modern science, but it is nevertheless a revolutionary theory in that its acceptance commits us to a radical reconstruction of our most fundamental physical concepts. In attempting this task of reconstruction it is essential that we should understand the reasons which led to the formation of the older concepts, in order that, if these are eventually discarded, we may ensure that nothing of value is lost.

Perhaps the most important lesson of the history of science is the abiding value of the result of a physical experiment carefully and accurately carried out under definite conditions. The chapter of scientific history with which we shall be most concerned is very largely the story of an always-changing theory based upon a growing body of unchanging experimental facts.

The problem of relativity appears for the first time in modern science in the work of Newton. The achievements of Galileo, Kepler, and Huygens were gathered up into a comprehensive theory of motion. It is clear that before such a theory can be formulated we must have a definition of motion, so that an observer may at any rate decide whether a particular body is moving or not. Now it is a familiar fact that we can observe only the relative motions of bodies, and that we cannot observe how any particular body is moving without reference to other bodies. In order to meet this difficulty, Newton adopted the hypothesis of an absolute space. The velocity and acceleration of a body mean its velocity and acceleration relative to absolute space, and no plan has yet been revealed by which these can be measured. This logical defect lies at the root of Newton's theory, that the terms in which it is expressed cannot be defined in such a way that they are unambiguously susceptible to physical measurement. Newton himself saw this difficulty very clearly, and he certainly would not have passed it by if it did not seem to him that there was a solution. The "fixed" stars are outside the solar system, and apparently unaffected by its motions. They might be used to define absolute space. This was the solution which Newton adopted, although with truly prophetic foresight he admitted "it may be that there is no body really at rest to which the places and motions of others may be referred." Subsequent discovery proved the wisdom of this reservation. The stars are not fixed even relatively one to another, but move with discordant and sometimes almost inconceivably great velocities. The logical difficulty returned with undiminished force, and it is clear that sooner or later it had to be faced and solved by some elaboration or reconstruction of the theory. However, the problem was left unanswered until our own day, partly because it was hoped that a solution would come from other branches of physics, but perhaps mainly because the stars are at such great distances that the ambiguity in the "absolute space" which they specify seemed unlikely to produce any measurable error in the application of the Newtonian laws to motions within the solar system. Thus it came about that the theory of relativity in its first form did not grow out of mechanics, but from other branches of physical knowledge—optics, electricity, and magnetism. We will endeavour to show how the fundamental ideas in these subjects gradually changed during the nineteenth century until they led to the formulation of the relativity theory. It will be convenient to begin by forming some idea of the state of knowledge in these branches of physics in the opening years of the nineteenth century. Thanks to the work of Newton and the way in which it had been pushed forward by the great French school of mathematicians, mechanics was very much what it is to-day, so that one may say roughly that the whole of the mechanics now required for our degree examinations was known. Lagrange and others had developed the Newtonian mechanics into a great and complete system which was thought to be capable of comprehending the whole of physics. Given the position and motion of all the bodies of the universe at any one instant, a master mathematician could work out the complete history of things past and future.

In optics the theory of what we now call geometrical optics was fairly well advanced, but practically nothing was known of physical optics. The laws of reflection and refraction and their application to the construction of lenses and telescopes; the phenomena of dispersion, but not the Fraunhofer lines in the solar spectrum; Newton's rings and the most elementary facts of polarisation,—would have been a fairly exhaustive syllabus in optics in the year 1800. Newton's corpuscular theory of light still held the field. It is true that in one form or another a wave theory had often been proposed, notably by Newton's contemporary Huygens, but as men thought always of a longitudinal wave, the facts of polarisation were held to be an insuperable barrier to such a theory. Rather curiously, the idea of an ether was already familiar in optics. In order to account for Newton's rings, Newton invented a theory under which his corpuscles suffered from fits of easy reflection and easy refraction, which were transmitted to the corpuscle through an ether filling all space. This same ether transmitted instantaneously the forces of gravitation between bodies. We have omitted to mention one isolated effect which was known and destined to play an important part in the evolution of later optical theories. On the corpuscular theory it is clear that, if a fixed star is observed by a moving telescope which has a component velocity v perpendicular to the direction of the star, then the telescope must not be pointed directly at the star, but at a point whose angular distance from the star is v/c, where c is the velocity of light. In consequence of the motion of the earth in its orbit round the sun, the stars will accordingly appear to describe small ellipses about their mean positions. This effect of stellar aberration had been observed and explained by Bradley in 1728.

When we turn to electricity and magnetism, we find that even less was known in the year 1800. The lodestone and permanent magnets made by its aid were used in navigation. In the reign of Elizabeth, Gilbert of Colchester had studied magnets, and had also discovered a large number of substances which became electrified on rubbing. Ten years before (1790) Galvani had constructed the first galvanic cell, and Volta's "pile" was the latest scientific novelty. Little was known of electrostatic induction or the properties of electric currents and nothing of any connection between electricity and magnetism. Although both a one-fluid and a two-fluid theory of electricity had been mooted, and Coulomb and Cavendish were laying the foundations for future advance by their quantitative investigation of the law of attraction, we may say that electricity and magnetism consisted of a few isolated and mysterious "effects."

The early advance of mechanics had an important effect which we can trace through the greater part of the century. As the knowledge of other branches of physics increased men tended to explain the new knowledge in terms of mechanical theories. Thus it was natural that when optics

demanded an eth
regarded as anoth
the laws of Newto

The opening o
marked a great a
Young revived tl
and in the foll
it by the series
which we now
ether assumed
was natural in
thought of as
same laws as, a
only in degree f
In order to
aberration whi
corpuscular th
suppose that tl
earth did not p
earth. Thus th
was originally
in its nature th
interstices of
through a grov

ssed through the
ained at rest. It
ut a difficulty in
theory, refraction
the velocity of
substance to its
. Accordingly,
the fixed ether
pend upon the
tested this by
n of light from
find any such
lty to Fresnel.
to Arago, met
cation in his
material body
ain fraction of
nely, $1 - 1/\mu^2$,
lex. Since the
y nearly unity,
ir is very small,
is very nearly
that the ether
other hand, in

glass of refractive
carried along wit
the velocity of th
was directly verifi
Fizeau in 1851.
an experiment
by Boscovich mo
fore. He observ
by means of a te
found that it
telescope was fi
foretold by Fres
had been expl
ging coefficient
results so well c
coefficient beca
had to be satis

In the mean
had been made
In 1819 Oerste
a magnet of a
rent. Ampère
to the discover
magnetic field c

nlarged our know-
ıd discovered the
ets upon electric
mathematical ex-
oère and Faraday.
ise on Electricity
ed an important
ıce. During the
ry a very scanty
electricity had
comprehensive
the influence of
attached great
the medium."
vas still formed
out of the two
ecome the more
c strains it pro-
orces, and by its
ted light. And
ether of Fresnel,
the motion of
ether satisfied
ntal results.

The rift in the lute appeared in 1881, when Michelson performed an experiment which was originally suggested by Maxwell. We shall have to examine this experiment in detail later, but for the present it is sufficient to note that it was an experiment designed to measure the relative velocity of the earth with respect to the surrounding ether.

According to Fresnel, the ether inside the earth is dragged, but the ether immediately outside, e.g. in a laboratory, is at rest,—at least to the approximation to which the refractive index of air is unity. The experiment was performed and the result was in direct contradiction to the predictions of Fresnel's theory. It appeared that there was no relative velocity as between the earth and the ether immediately outside the earth. It was clear that some modification of the theory was necessary and it may be well to recall the three experimental results which had to be satisfied by any proposed theory :—

(1) Stellar aberration of an amount which is independent of the medium inside the telescope.

(2) The increase of the velocity of light when it is propagated in the direction of motion of a moving medium (Fizeau's experiment).

(3) The null result of Michelson's experiment.

Some time earlier (1845) the mathematician Stokes had felt doubts as to Fresnel's ether on somewhat theoretical grounds. The dragging coefficient imposed a discontinuity in the motion of the ether at the surface of a moving body. In those days discontinuities were less in favour among physicists than they are now, and Stokes tried to remove the difficulty by supposing that the ether was a viscous fluid, so that the ether inside the earth is dragged along in accordance with Fresnel's coefficient, but that at the surface the velocity does not immediately fall to zero, but gradually and continuously falls off just as in the case of a sphere moving through a viscous fluid. The velocity of the ether immediately outside the earth is then approximately equal to the velocity of the earth and the result of Michelson's experiment, had it been known, would have suggested Stoke's theory. At first sight it would seem, however, that such an hypothesis would fail to account for aberration. Stokes showed that this was not so,—that his theory would give the correct aberration so long as the motion of the ether was of the type which is called irrotational in hydrodynamics.

This brought to light a new difficulty, for it appears that there is no possible irrotational motion of a fluid surrounding a moving sphere such that the fluid is at rest at infinity and there is no slip of the fluid over the surface of the sphere. At least, such a motion is impossible if the fluid is incompressible. The analysis of the motion of a compressible viscous fluid is a very difficult problem of which very little is known even to-day. Thus one method of adjusting Fresnel's ether to the result of Michelson's experiment was disposed of in advance. This possibility has, however, been returned to in recent years, and Planck has shown that if the ether is compressible we can make the slip at the earth's surface as small as we please, provided that there is a condensation of ether round the earth. But in order to reduce the slip to 1 per cent., when it would be too small to be measured by Michelson's experiment, the density of ether immediately outside the earth is about 80,000 times its density at a great distance. Yet this enormous change in the density of the ether produces no measurable difference in the properties of the ether or in the propagation of light. There seems little hope of progress in this direction.

The Michelson experiment was first explained by an *ad hoc* hypothesis suggested independently by Fitzgerald and Lorentz—that a material body moving with velocity v through the ether was contracted in the ratio $1 : \sqrt{(1 - v^2/c^2)}$, where c is the velocity of light. This hypothesis explained the single result which it was designed to explain, but no independent evidence of the existence of the contraction could be obtained. In fact, there were certain difficulties about the conception of the contraction—what, for example, happens when we rotate a wheel at high speed: is its circumference contracted without change in its diameter? However, the Fitzgerald-Lorentz hypothesis might have remained were it not for a brilliant theoretical development by Lorentz himself. In order to understand this, we must return to the consideration of the development of the electromagnetic theory. Maxwell's theory was really the analytical expression of two physical laws:—

(1) *The Law of Faraday.*—The integral of the electric force round any circuit is proportional to the rate of change of the flux of magnetic induction through that circuit.

(2) *The Law of Oersted as Amended by Maxwell.*—The line integral of the magnetic force round any circuit is proportional to

the total flow of current through that circuit. The total current includes the displacement current which is the rate of change of the electric induction.

Four quantities play an important part in these laws—the electric force E, the electric induction D, the magnetic force H, the magnetic induction B. Maxwell assumed that these are connected by the empirical relations

$$B = \mu H, \; D = \epsilon E$$

where μ, ϵ are the magnetic permeability and the specific inductive capacity.

In view of Fresnel's work it was natural that the question should arise as to how Maxwell's equations were to be applied to moving media. Strangely enough, Maxwell does not appear to have considered this problem in any detail. Hertz, however, took Maxwell's equations and assumed that when applied to a moving medium the circuits referred to above are to be interpreted as circuits fixed in the medium, while the relations $B = \mu H$ and $D = \epsilon E$ are maintained. These assumptions lead to the conclusion that Fresnel's dragging coefficient would be unity, and thus Hertz's theory of moving media was in direct conflict with experiment.

Lorentz attacked the problem from an entirely new point of view by examining the basis of the relations $B = \mu H$ and $D = \epsilon E$. He assumed that the differences between B and H and between D and E were given by $D = E + P$, $B = H + M$, where P is the electric polarisation and M the magnetisation of the medium. P and M he regarded as due to the influence of E and H upon the motion of the electrons contained in the atoms of the material. This development of the electron theory of matter proved most fruitful. A number of hitherto unexplained optical effects were accounted for and the way was prepared for a theory of moving media. The electric induction which plays the part of the electric force inside matter consists of two parts—(1) E residing in the ether and unaffected by the motion of the medium; (2) P arising from the electrons of the medium and intimately bound up with the motion of the medium.

There is no need for us now to follow all the intricacies of Lorentz's theory. It is sufficient to note that he developed a complete theory of moving media based upon (*a*) Maxwell's equations for free space; (*b*) his own hypothesis as to the relation between E and D and between B and H; (*c*) the Fitzgerald-Lorentz contraction. The equations which expressed this theory were naturally more complicated than Maxwell's equations, but Lorentz showed that, by introducing a new variable τ in place of the time t, the equations for a moving medium took the same form as the equations for free space. Lorentz called τ the "proper time," but he regarded it as no more than a mathematical variable which facilitated the solution of the problem of moving media. Einstein carried the process further by a bold step. Since only relative motions can be observed how can we say whether our medium is moving or not, and how can we distinguish between the fictitious mathematical "proper time" τ and the absolute time t? Einstein assumed that the proper time τ was the time measured by physical observation, and that, therefore, the equations for a "moving" medium were in relation to the time observed in that medium the same as if the medium were at rest. On this he based his principle of relativity that the laws of nature are such that no experiment can reveal an absolute velocity, or, what comes to the same thing, a velocity relative to the ether.

We shall have to examine the work of Lorentz and Einstein in greater detail, but this brief sketch may serve to show how their work falls into place in a continuous attempt to build up a theory of the ether which shall conform to the results of physical experiment.

II

THE MICHELSON AND MORLEY EXPERIMENT AND THE LORENTZ TRANSFORMATION

IN its essence this experiment was a comparison of the velocity of propagation of light in two mutually perpendicular directions. A ray of light OA is incident at an angle of 45° on a half-silvered mirror, so that the reflected and transmitted rays are perpendicular. These travel along paths AB, AC respectively, which in the ideal case may be supposed to be exactly equal in length. They are incident normally upon plane mirrors at B and C, and are reflected back along their respective paths, so that both rays arrive again at A. The transmitted part of the ray originally reflected, and the reflected part of the ray originally transmitted, will then be superposed along AP in a direction which is perpendicular to the direction of the original ray OA. If the velocity of light is the same in the directions AB, AC,

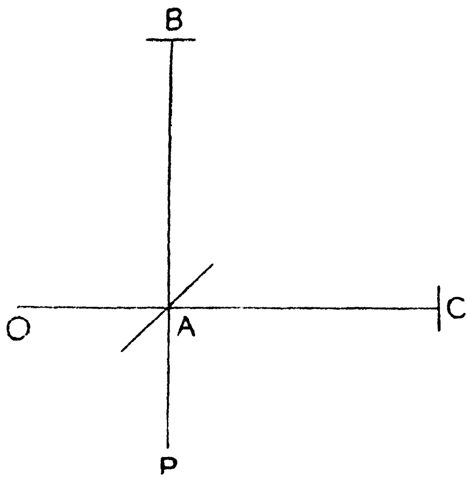

FIG. 1.

the two rays superposed along AP will be in phase; but if there is a difference of velocity in the two directions there will be a consequent difference of phase between the two rays in AP and this will be made manifest by interference. This is, of course, a very much idealized account of a highly technical experiment, but it contains the essential principles.

In order to keep our ideas as definite as possible, we will interpret this experiment on the basis of Fresnel's fixed ether hypothesis. If the whole apparatus is at rest in the ether, we should expect the velocities along AC and AB to be the same. If, however, the whole apparatus is moving through the ether, say with velocity v in the direction AC, then it will appear that the time of passage along ACA is greater than along ABA by an amount which will depend upon v, and which may be measured by the interference of the two rays superposed along AP. Thus, on the assumption that light is propagated in the ether with the same velocity in all directions, the experiment provides a means of measuring the velocity of the apparatus through the ether.

Assuming that the sun is at rest in the ether, the earth, owing to its annual motion round the sun, has a velocity of abo 18 miles per second, while owing to the diurnal rotation a point on the earth's equator has a velocity of about one-third of a mile a second. If the experiment is performed in a laboratory the apparatus is, according to Fresnel's theory, moving through the ether with a speed of 18 miles per second, and the delicacy of Michelson's experiment was such that a velocity of this order could be detected. The experiment failed to produce any evidence of this or any other velocity through the ether. The problem is not materially changed if we admit the possibility of a motion of the sun through the ether. By applying a process of averaging to the observed motions of the "fixed" stars, astronomers have arrived at the conclusion that the whole solar system is moving through space with a velocity of about 10 miles per second. It is true that this might at a particular time reduce the velocity of the earth through the ether to 8 miles per second, but, on the other hand,

..x months later it would increase it to 28 miles per second. The only way in which we can suppose that our laboratory is *permanently* at rest in a fixed ether is to undo the work of Copernicus and Kepler, and to return to a Ptolemaic theory of the universe. If, on the other hand, we admit a motion of the earth through the ether, we must suppose that we have left out of account some compensating influence which prevents Michelson's experiment from detecting that motion. Such a compensating influence was proposed by Fitzgerald and Lorentz in their famous contraction hypothesis: a body moving through the ether undergoes a contraction of length in the direction of its motion. Thus in Michelson's experiment the path ACA, which, owing to the motion through the ether, would correspond to the longer time of passage, is contracted in length by just such an amount that the time of passage is the same for the two paths. Such a contraction would not be revealed by our ordinary measurements, since presumably our measuring scales are also contracted in the same ratio. The Fitzgerald-Lorentz hypothesis introduced a distinction between a measured length and a real length.

We will follow out the implications of the result of Michelson's experiment, and it will

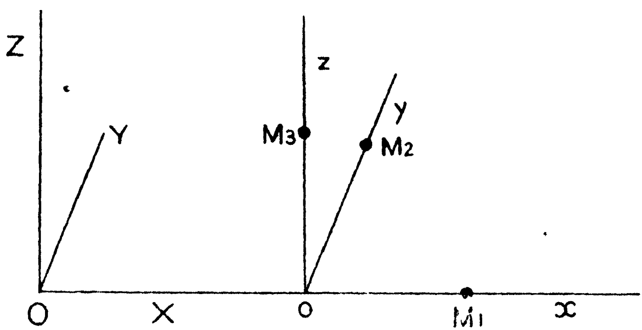

FIG. 2.

help to keep our ideas clear if we adhere to Fresnel's hypothesis of a fixed ether, while admitting the distinction between real and measured quantities. We will return later to a discussion of the meaning of this distinction.

Suppose that a set of axes $o(x, y, z)$ are drawn fixed in our laboratory. The coordinates x, y, z are measured lengths, and we also have a means of determining a measured time t. We admit that we are moving through the ether with an unknown velocity, and suppose that this is constant and equal to v in the direction ox. We take a set of axes O(XYZ) fixed in the ether. Since the difference between real and measured lengths is due to motion through the ether, we may suppose that X, Y, Z are real lengths, and further that, corresponding to the measured time t, there is a real time T. The two sets of axes may be taken to coincide at time $t = 0$. By Michelson's experiment, we find that the measured velocity of light relatively to our apparatus is the same in all directions, and our units may be adjusted so that this measured velocity is the same as the real velocity c. Let mirrors M_1, M_2, M_3 be placed on the axes at equal measured distances l_1, l_2, l_3 $(= l)$ from o, and at time $t = 0$ let a pulse of light be emitted from o and return to that point after reflection at the three mirrors.

Consider first the ray which passes alon *ox*; let it reach M_1 at measured time t_1 return to *o* at time t_1''. Then

$$t_1' = \frac{l}{c}, \quad t_1'' = \frac{2l}{c}.$$

If the corresponding real quantities are denoted by capital letters, and we note that the velocity with which the light approaches M_1 is $c - v$, while that with which it approaches *o* on its return journey is $c + v$, we have

$$T_1' = \frac{L_1}{c - v},$$

$$T_1'' = \frac{L_1}{c - v} + \frac{L_1}{c + v} = \frac{2cL_1}{c^2 - v^2}.$$

Next consider the ray which passes along *oy*. With a similar notation for measured time, we have

$$t_2' = \frac{l}{c}, \quad t_2'' = \frac{2l}{c}.$$

The real path of this ray is the hypotenuse of a right-angled triangle of sides L_2 and vT_2'. Hence $c^2T_2'^2 = L_2^2 + v^2T_2'^2$, or

$$T_2' = \frac{L_2}{\sqrt{(c^2 - v^2)}}, \quad T_2'' = \frac{2L_2}{\sqrt{(c^2 - v^2)}}.$$

.le experimental result is that the two
ays arrive back at o at the same time. Hence $T_2'' = T_1''$, or

$$\beta L_1 = L_2,$$

where $\beta = 1/\sqrt{(1 - v^2/c^2)}$, and is therefore a fraction greater than unity.

The original assumption of the Fitzgerald-Lorentz hypothesis was that the dimensions of a body in a direction perpendicular to the direction of motion are unchanged, or, in other words, measured and real lengths are the same in any direction perpendicular to that of the motion through the ether. It has been shown that no material increase of generality is obtained by abandoning this assumption. Hence we may take $L_2 = l_2 = l_1$, and then

$$L_1 = l_1/\beta \quad . \quad . \quad . \quad . \quad (1)$$

Further, the real time T_1 of the time of the double passage along either path is related to the corresponding measured time by

$$T_1'' = \beta t_1'' \quad . \quad . \quad . \quad (2)$$

These results may be expressed as relations between the *measured* co-ordinates x, y, z with respect to o and the *real* co-ordinates X, Y, Z with respect to O. For the *real* co-ordinates of x, y, z with respect to o are $x/\beta, y, z$, which are respectively equal to $X - vT$, Y, Z. Hence

$$x = \beta(X - vT), y = Y, z = Z.$$

The relation between the measured time t and the real time T may be obtained in a similar way, but the argument is clearer if we note that the wave surface of a pulse of light emitted from the origin at time $t = 0$ is a sphere with centre o in measured lengths and times, whereas it is a sphere with centre O in real lengths and times. That is to say, that the following two equations are equivalent :—

$$x^2 + y^2 + z^2 = c^2t^2, X^2 + Y^2 + Z^2 = c^2T^2$$

The first gives

$$\beta^2(X - vT)^2 + Y^2 + Z^2 = c^2t^2$$

and using the second we may solv

terms of T and X. We thus obtain the famous Lorentz transformation

$$\left.\begin{array}{c} x = \beta(\mathrm{X} - v\mathrm{T}), \quad y = \mathrm{Y}, \quad z = \mathrm{Z}, \\ t = \beta\left(\mathrm{T} - \dfrac{v\mathrm{X}}{c^2}\right). \end{array}\right\} \quad (3)$$

From our present point of view these represent the relations between our measured lengths and times and the corresponding real lengths and times measured with respect to the fixed ether.

Perhaps the greatest difficulty which has been felt by many in approaching the new theory is that x, t each depends upon both X and T, so that space and time appear to be "mixed up." This difficulty will disappear if we are careful to see exactly what is implied by these relations. If two events take place at the same point in the ether they have the same X. The first of the orentz relations then asserts that they will occur at the same place in our laboratory they occur at the same time T. This ›usly true if in fact we are moving through the ether. The fourth of the Lorentz relations is not quite so easy to dispose of. It asserts that if two events occur at the same real time T, *i.e.* if they are really simultaneous, they will not be simultaneous in our measured time unless they occur at the same place in the ether (or at least have the same X). This contradicts our usual assumption that we can determine the simultaneity of events with certainty; that, for example, we can synchronize two distant clocks. A little reflection, however, will show the great difficulty of suggesting any means by which this may be done without knowing our velocity through the ether. The synchronization of clocks is a practical problem, and two methods have been largely used by astronomers. Portable clocks are compared in turn with the two clocks to be synchronized, but in order to do this the portable clocks must move through the ether. Their parts will be subject to the Fitzgerald-Lorentz contraction and to the order of accuracy with which we are now dealing it would be bold to predict what would happen during the course of their journey. The more modern method is by means of wireless signals, and to be exact we must correct for the time taken to propagate the signals. These, like light waves, are propagated with constant velocity through the ether. If both our clocks are moving through the ether the correction will depend upon their common velocity. For example, if our clocks are at o and M_1 in Fig. 2, and the signal is sent from the first to the second, the correction for the time of propagation would be

$$\frac{L_1}{c-v} = \frac{l_1}{\beta(c-v)}$$

and this correction cannot be made unless v is known. We can, of course, make the clocks synchronous in measured time by using the experimental result that the velocity of light in measured lengths and times is the same in all directions, but clocks so synchronized will not be synchronous in the real time T.

Corresponding to measured lengths and distances there will be measured velocities which will in general be different from the true velocities. From equations (3) we have

$$\delta x = \beta(\delta X - v\delta T), \quad \delta y = \delta Y, \quad \delta z = \delta Z,$$
$$\delta t = \beta\left(\delta T - \frac{v}{c^2}\,\delta X\right).$$

If the measured velocities of a point are given by $u_x = \delta x/\delta t$ . . . and the corresponding true velocities by $U_x = \delta X/\delta T$ . . ., we have

$$\left.\begin{aligned} u_x = \frac{U_x - v}{1 - \dfrac{vU_x}{c^2}}, \qquad u_y &= \frac{U_y}{\beta\left(1 - \dfrac{vU_x}{c^2}\right)}, \\ u_z &= \frac{U_z}{\beta\left(1 - \dfrac{vU_x}{c^2}\right)} \end{aligned}\right\} \quad (4)$$

In a similar way we can obtain the relations between the measured and true accelerations of a moving point. If $f_x = \delta u_x/\delta t$ . . . , and $F_x = \delta U_x/\delta T$ . . . ,

$$\left.\begin{aligned} f_x &= \frac{F_x}{\beta^3\left(1 - \frac{vU_x}{c^2}\right)^3} \\ f_y &= \frac{F_y}{\beta^2\left(1 - \frac{vU_x}{c^2}\right)^2} + \frac{vU_yF_x}{\beta^2c^2\left(1 - \frac{vU_x}{c^2}\right)^3} \\ f_z &= \frac{F_z}{\beta^2\left(1 - \frac{vU_x}{c^2}\right)^2} + \frac{vU_zF_x}{\beta^2c^2\left(1 - \frac{vU_x}{c^2}\right)^3} \end{aligned}\right\} \quad (5)$$

The formulæ (4) and (5) may be used as the basis of a complete theory of the kinematics of measured motion, but we will note only some of the simpler consequences of (4). If the true velocity of a point is equal to the velocity of light, say $U_x = c$, $U_y = U_z = 0$, we have $u_x = c$, $u_y = u_z = 0$. More generally, if the true velocity of a int is in any direction, but is less than elocity of light, then the measured also less than that of light.

Again, if $u_x = u_y = u_z = 0$, we have $U_x = v$, $U_y = U_z = 0$, or v is the true velocity through the ether of any point fixed with respect to the axes $o(x, y, z)$. From this point of view it might more properly have been denoted by V. The inconsistency is, however, removed if we note that when $U_x = U_y = U_z = 0$, $u_x = -v$, $u_y = u_z = 0$, so that v is also numerically equal to the measured velocity of the ether with respect to the axes $o(x, y, z)$, assuming that any means could be found by which it could be measured.

Finally, we note a very important property of the Lorentz transformation. Equations (3) express the measured co-ordinates in terms of the real co-ordinates. If they are solved for the latter, we obtain

$$\left.\begin{aligned} X = \beta(x + vt), \quad Y = y, \quad Z = z \\ T = \beta\left(t + \frac{vx}{c^2}\right) \end{aligned}\right\} \quad (3')$$

Allowing for the fact established above, that the measured velocity of the ether with respect to o is equal and opposite to the real velocity of o with respect to the ether, we see that the relations between the real co-ordinates and the measured co-ordinates are completely reciprocal. This is the point at which we begin to suspect the reality of the real co-ordinates.

III

THE LAWS OF MOTION AND ELECTROMAGNETISM

WE will retain our distinction between the real lengths and times measured with respect to the fixed ether and the lengths and times actually measured in a laboratory moving through the ether. We have ascertained the relations between the measured lengths and times, and the corresponding real lengths and times which are dictated by the result of the Michelson-Morley experiment, and we will now proceed to examine the relations between the real and measured values of other fundamental physical quantities. Prominent among these are mass and force in terms of which Newton's laws are expressed. We assume that Newton's laws are true with reference to a set of axes fixed in the ether, *i.e.* our axes of X, Y, Z, making, however, this extension, prompted by the results of experiment upon bodies moving with high speeds—that the mass is not a constant but is a function of the speed. This at once leads to a difficulty, for we are accustomed to express Newton's laws in two forms, which are equivalent only so long as the mass is constant—the law of mass acceleration,

$$P = MF,$$

and the law of momentum,

$$P = \frac{d}{dT}(MU).$$

If M is a function of U, and therefore of T, the two forms are no longer equivalent. We adopt the second. This may be expressed as a law of mass acceleration, but if this is done the mass of a particle is different for forces in the direction of motion from what it is for forces at right angles to that direction. If the particle is moving in the direction of X with velocity v and accelerations F_x, F_y, F_z, we have

$$P_x = \frac{d}{dT}(MU) = \left(M + v\frac{dM}{dv}\right)F_x$$

while

$$P_y = \frac{d}{dT}(MV) = MF_y.$$

M is spoken of as the "transverse" mass, while the "longitudinal" mass is given by

$$M_l = M + v\frac{dM}{dv}.$$

It should be noted that if we use the momentum form for the laws of motion, the mass is the same for all directions of applied force, and is equal to the transverse mass.

Suppose that a particle of mass m is instantaneously at rest in the measured co-ordinates, but has accelerations f_x, f_y, f_z. The measured forces will be given by

$$P_x = mf_x,\ P_y = mf_y.$$

By formulæ (5) of the last lecture, when $U_x = v,\ U_y = o,\ U_z = o,$

$$f_x = \beta^3 F_x,\ f_y = \beta^2 F_y.$$

Assume that the measured units of force and mass are so chosen that the measured force is equal to the true force in the direction of motion. Then

$$P_x = M_l F_x = \frac{M_l}{\beta^3} f_x = \frac{M_l}{m\beta^3} p_x.$$

Hence

$$M_l = m\beta^3,$$

or

$$M + v\frac{dM}{dv} = m\beta^3.$$

This gives

$$vM = m\int\frac{dv}{(1 - v^2/c^2)^{\frac{3}{2}}} = \frac{mv}{(1 - v^2/c^2)^{\frac{1}{2}}} + A.$$

Since $M = m$ when $v = 0$, we have for the transverse mass

$$M = \beta m.$$

The longitudinal mass is then given by

$$M_l = m\left(\beta + v\frac{d\beta}{dv}\right) = m\beta^3.$$

This result has an important consequence. The rate at which work is done by the forces is

$$vP_x = v\beta^3 m\,\frac{dv}{dT} = \frac{d}{dT}\left(\frac{mc^2}{\sqrt{1 - v^2/c^2}}\right).$$

d1

The quantity within brackets is such that its rate of increase is equal to the work done by the forces. We may clearly add any constant without affecting this result, and choosing this constant so that the expression vanishes with v we have

$$T = m\left\{\frac{c^2}{\sqrt{1 - v^2/c^2}} - c^2\right\}.$$

This is the function which plays the part of the kinetic energy when the mass depends upon the velocity. Note that if we neglect the fourth and higher powers of v/c, $T = \frac{1}{2}mv^2$.

Returning to the relations between the measured and true forces, we have for a particle at rest in the measured system,

$$p_y = mf_y = m\beta^2 F_y = \beta M F_y = \beta P_y.$$

Hence

$$p_x = P_x,\ p_y = \beta P_y,\ p_z = \beta P_z.$$

Following the same line of argument, we find for a particle *at rest in the true co-ordinates*

$$P_x = p_x,\ P_y = \beta p_y,\ P_z = \beta p_z.$$

The equations of the electromagnetic field as adopted by Lorentz are

$$-\frac{1}{c}\frac{\partial H_x}{\partial T} = \frac{\partial E_z}{\partial Y} - \frac{\partial E_y}{\partial Z}$$

$$\frac{1}{c}\left(\frac{\partial E_x}{\partial T} + PU_x\right) = \frac{\partial H_z}{\partial Y} - \frac{\partial H_y}{\partial Z}$$

$$-\frac{1}{c}\frac{\partial H_y}{\partial T} = \frac{\partial E_x}{\partial Z} - \frac{\partial E_z}{\partial X}$$

$$\frac{1}{c}\left(\frac{\partial E_y}{\partial T} + PU_y\right) = \frac{\partial H_x}{\partial Z} - \frac{\partial H_z}{\partial X}$$

$$-\frac{1}{c}\frac{\partial H_z}{\partial T} = \frac{\partial E_y}{\partial X} - \frac{\partial E_x}{\partial Y}$$

$$\frac{1}{c}\left(\frac{\partial E_z}{\partial T} + PU_z\right) = \frac{\partial H_y}{\partial X} - \frac{\partial H_x}{\partial Y}$$

$$\frac{\partial H_x}{\partial X} + \frac{\partial H_y}{\partial Y} + \frac{\partial H_z}{\partial Z} = 0$$

$$\frac{\partial E_x}{\partial X} + \frac{\partial E_y}{\partial Y} + \frac{\partial E_z}{\partial Z} = P,$$

while the force acting on a moving charge e is P_x, P_y, P_z, where

$$P_x = e\left(E_x + \frac{U_y}{c}H_z - \frac{U_z}{c}H_y\right)$$

$$P_y = e\left(E_y + \frac{U_z}{c}H_x - \frac{U_x}{c}H_z\right)$$

$$P_z = e\left(E_z + \frac{U_x}{c}H_y - \frac{U_y}{c}H_x\right).$$

In these equations E_x, E_y, E_z and H_x, H_y, H_z are respectively the components of electric and magnetic force; P is the density of charge, and U_x, U_y, U_z are the components of the velocity of the charge. There are two things to notice about these equations. Firstly, the units employed are not those belonging to either of the systems commonly in use. A unit charge is defined to be such that two units at a distance apart of 1 cm. repel each other with a force of $1/4\pi$ dynes; it is thus smaller than the electrostatic unit in the ratio $1 : \sqrt{(4\pi)}$. The second set of three equations expresses the law that the line integral of H round any circuit is equal to $\frac{1}{c} \times$ total current through the circuit. Allowing for the change already made in the unit of charge, it follows that H is measured in a unit which is greater than the electromagnetic unit in the ratio $\sqrt{(4\pi)} : 1$. These units are particularly convenient in theoretical work, as they make the equations symmetrical and avoid the frequent occurrence of the factor 4π. Secondly, it will be noted that the equations ignore the distinction between the electric force and induction and between the magnetic force and induction. On Maxwell's view the equations would therefore be those applicable to free space unoccupied by matter. Lorentz, however, assumed that the above equations are strictly true everywhere, even in the interior of an electron, and he and Minkowski showed that the difference between the force and induction in each case could be explained as due to interatomic electronic motion. It can be shown that the above microscopic equations, when averaged over a volume sufficiently large to contain many electrons, lead to a set of macroscopic equations involving the electric and magnetic inductions, which differ from the corresponding forces by terms depending upon the polarization of the medium, *i.e.* the distribution and motion of the concealed electronic charges. Finally, these equations are supposed to hold for a set of axes fixed in the ether, and we have accordingly written them in terms of X, Y, Z, T.

We will find the equations between the corresponding measured quantities on the assumption that the measured value of an electric charge at rest in the measured system is equal to its true value. Defining the measured electric force as the force on a unit charge at rest in the measured coordinates, we have

$$e_x = p_x = P_x = E_x.$$

Similarly,

$$e_y = p_y = \beta P_y = \beta\left(E_y - \frac{v}{c} H_z\right).$$

Thus

$$e_x = E_x, \qquad e_y = \beta\left(E_y - \frac{v}{c} H_z\right),$$

$$e_z = \beta\left(E_z + \frac{v}{c} H_y\right).$$

From (3) of the last lecture

$$\frac{\partial}{\partial X} = \beta \frac{\partial}{\partial x} - \frac{\beta v}{c^2} \frac{\partial}{\partial t}, \quad \frac{\partial}{\partial Y} = \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial Z} = \frac{\partial}{\partial z},$$

$$\frac{\partial}{\partial T} = -\beta v \frac{\partial}{\partial x} + \beta \frac{\partial}{\partial t}.$$

Substituting in Maxwell's equations, as set out above, we find that they remain completely unchanged save that the true quantities X, T, E, H, U . . . , are replaced by the corresponding measured quantities, provided that

$$h_x = H_x, \qquad h_y = \beta\left(H_y + \frac{v}{c}E_z\right),$$

$$h_z = \beta\left(H_z - \frac{v}{c}E_y\right)$$

$$\rho = \beta\left(1 - \frac{vU_x}{c^2}\right)P.$$

The last relation is consistent with the assumption that the measured value of a charge at rest in the measured co-ordinates shall be the same as its true value, for in this case $U_x = v$ and $\rho = P/\beta$, as it should be if the measured volume of the element is greater than its true volume in the ratio $\beta : 1$.

The method by which these results are established will be sufficiently illustrated if we consider the case of the fourth and last of the electromagnetic equations. The fourth gives

$$\frac{1}{c}\left(-\beta v\frac{\partial E_x}{\partial x}+\beta\frac{\partial E_x}{\partial t}+PU_x\right)=\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}.$$

The last gives

$$\beta\frac{\partial E_x}{\partial x}-\frac{\beta v}{c^2}\frac{\partial E_x}{\partial t}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}=P.$$

Hence

$$\frac{1}{c}\left(\beta\frac{\partial E_x}{\partial t}-\frac{\beta v^2}{c^2}\frac{\partial E_x}{\partial t}+P(U_x-v)\right)$$
$$=\frac{\partial}{\partial y}\left(H_z-\frac{v}{c}E_y\right)-\frac{\partial}{\partial z}\left(H_y+\frac{v}{c}E_z\right),$$

or

$$\frac{1}{c}\left(\frac{\partial e_x}{\partial t}+\rho u_x\right)=\frac{\partial h_z}{\partial y}-\frac{\partial h_y}{\partial z}$$

since

$$\rho u_x=\beta\left(1-\frac{vU_x}{c^2}\right)\frac{(U_x-v)P}{1-\dfrac{vU_x}{c^2}}$$
$$=\beta(U_x-v)P.$$

The other equations transform in a similar way so that the measured quantities obey equations in the measured co-ordinates, which are of precisely the same form as the equations obeyed by the real quantities in the real co-ordinates, provided that

$$e_x = \mathrm{E}_x, \qquad e_y = \beta\left(\mathrm{E}_y - \frac{v}{c}\mathrm{H}_z\right),$$

$$e_z = \beta\left(\mathrm{E}_z + \frac{v}{c}\mathrm{H}_y\right)$$

$$h_x = \mathrm{H}_x, \qquad h_y = \beta\left(\mathrm{H}_y + \frac{v}{c}\mathrm{E}_z\right),$$

$$h_z = \beta\left(\mathrm{H}_z - \frac{v}{c}\mathrm{E}_y\right)$$

$$\rho = \beta\left(1 - \frac{v\mathrm{U}_x}{c^2}\right)\mathrm{P}.$$

These relations may be solved so as to express the true quantities in terms of the measured quantities. Thus

$$\mathrm{E}_x = e_x, \qquad \mathrm{E}_y = \beta\left(e_y + \frac{v}{c}h_z\right),$$

$$\mathrm{E}_z = \beta\left(e_z - \frac{v}{c}h_y\right), \text{ etc.}$$

IV

THE RESTRICTED PRINCIPLE OF RELATIVITY AND SOME CONSEQUENCES.

UP to this point we have adhered to the hypothesis of a fixed ether through which our laboratory is supposed to move with a velocity which is definite, although so far no way has been discovered by which it can be measured. In order to account for the result of Michelson's experiment, we have been led to admit a distinction between the actually measured values of physical quantities and their true values as measured with respect to the fixed ether. Assuming that the fundamental laws of motion and electromagnetism are true with respect to axes fixed in the ether, we have found the laws governing the corresponding quantities as measured with respect to axes moving through the ether. On certain assumptions, some of which may be avoided by a more exhaustive analysis, we have found that the physical laws for measured quantities are of precisely the same form as the laws for the corresponding true quantities. In other words, the velocity of motion through the ether does not appear in the equations for the measured quantities. This result corresponds to the negative fact that so far no physical measurement has been found which can determine the velocity of motion with respect to the fixed ether.

Further, we have found a complete reciprocity in the relations between true and measured quantities, so that an equation expressing a measured quantity in terms of the corresponding true quantity can be turned into one expressing the true in terms of the measured quantity merely by changing the sign of v. In order to see the significance of this reciprocity, let us regard our moving axes as fixed in the ether. Our measured quantities then become true quantities. The axes which were formerly regarded as being fixed in the ether are now moving with velocity $-v$. If we now enquire what will be the values of the measured quantities for these axes, we shall obtain the values which we have hitherto regarded as the true values. The distinction between the true value of a physical quantity and its measured value—a distinction which must seem unsatisfactory to the physicist—now disappears. Both sets of quantities are measured, but measured with respect to different sets of axes moving with different velocities through the ether.

We may take any set of axes *moving with a uniform velocity* through the ether, and regard these as the fixed axes of Newton and Maxwell. If we take a second set of axes moving uniformly with respect to the first, the physical quantities for the two sets of axes will be related by the laws given above, but the physical laws will be the same for both sets of axes. If at any time it is convenient, we may regard the second set of axes as fixed and the first as moving. This is the restricted principle of relativity.

We will now examine some of the consequences of this point of view, in order to show that the principle accounts for the governing experimental results. From the line of development it is clear that it accounts for the nul result of the Michelson experiment.

Stellar Aberration

Suppose plane waves of light are received from a distant star, fixed with respect to the axes of X, Y, Z, in a direction making an angle θ with the axis of X, and in the plane of XY. The light disturbance is of the form

$$f(\mathrm{X}\cos\theta + \mathrm{Y}\sin\theta + c\mathrm{T}).$$

With respect to axes x, y, z moving relatively to X . Y . Z with velocity v in the direction of x,

$$X \cos\theta + Y \sin\theta + cT$$

$$= \beta(x + vt) \cos\theta + y \sin\theta + \beta c\left(t + \frac{vx}{c^2}\right)$$

$$= x\left\{\beta\left(\cos\theta + \frac{v}{c}\right)\right\} + y \sin\theta$$

$$+ ct\left\{\beta\left(1 + \frac{v}{c}\cos\theta\right)\right\}$$

$$= \beta\left(1 + \frac{v}{c}\cos\theta\right)\{x\cos\theta' + y\sin\theta' + ct\},$$

where $\cos\theta' = \dfrac{c\cos\theta + v}{c + v\cos\theta}$, $\sin\theta' = \dfrac{c\sin\theta}{c + v\cos\theta}$.

To an observer moving and measuring with the axes x, y, z, the light is received at the angle θ', where

$$\sin(\theta - \theta') = \frac{v\sin\theta}{c + v\cos\theta},$$

or, neglecting squares and higher powers of v/c, the star is apparently deflected through an angle

$$\frac{v}{c}\sin\theta.$$

It should be noted that the symbol v now has a precise physical significance, namely, the velocity of the observer relative to the star. Consistent with this velocity, we may suppose that either the star or the observer is at rest in the ether; the result is the same in both cases.

The Doppler Effect

As a particular case of the above result, suppose that the light disturbance is of the form

$$\cos \frac{2\pi}{\lambda} (X \cos \theta + Y \sin \theta + cT),$$

so that to an observer at rest relatively to the star the light is of frequency

$$\nu = \frac{c}{\lambda}.$$

In the co-ordinates x, y, z the disturbance becomes

$$\cos \frac{2\pi}{\lambda'} (x \cos \theta' + y \sin \theta' + ct)$$

where

$$\frac{\lambda}{\lambda'} = \frac{\nu'}{\nu} = \beta \left(1 + \frac{v}{c} \cos \theta\right).$$

Apart from terms of the second order, this shows that there is an increase in the frequency of the light from a star given by $(\nu' - \nu)/\nu = \frac{1}{c}$ relative velocity of recession of the star in the line of sight.

Fresnel's Dragging Coefficient

The complete investigation of this problem demands an examination of the electric polarization of the medium, but it is possible to get a certain amount of information as follows. Consider a medium moving with a constant velocity v in the direction X relatively to an observer whose axes are X, Y, Z, and let x, y, z be axes fixed in the medium. Light is propagated through the medium with a velocity c/μ where μ is the refractive index. If the propagation is in the direction of motion, we have $u_x = c/\mu$ and

$$U_x = \frac{u_x + v}{1 + vu_x/c^2} = \frac{c + \mu v}{\mu(1 + v/\mu c)}.$$

Expanding in powers of v/c and neglecting squares and higher powers, this gives

$$U_x = \frac{c}{\mu} + v\left(1 - \frac{1}{\mu^2}\right).$$

Thus the velocity of propagation is increased by the amount $v\left(1 - \frac{1}{\mu^2}\right)$ as is required by Fresnel's formula.

THE FIELD OF A MOVING ELECTRON

Suppose that, as measured in a system of co-ordinates relatively to which it is at rest, the electron is built up of spherical layers of constant charge-density so that the electron and its field are symmetrical about its centre. Let the electron be at rest at the origin of the co-ordinates x, y, z, and examine the field as measured in the co-ordinates, X, Y, Z as defined above. In the latter co-ordinates the electron moves with uniform velocity v in the direction of X.

Since the electron is symmetrical and at rest in the co-ordinates x, y, z, its field is given by

$$e_x = \frac{x}{r}\phi, \quad e_y = \frac{y}{r}\phi, \quad e_z = \frac{z}{r}\phi,$$

$$h_x = h_y = h_z = 0,$$

where ϕ is the radial electric force and is a function of r only, where $r^2 = x^2 + y^2 + z^2$.

Employing the formulæ of transformation given in the last lecture, we have for the field, as measured in the co-ordinates X, Y, Z,

$$E_x = \frac{x}{r}\phi, \quad E_y = \beta\frac{y}{r}\phi, \qquad E_z = \beta\frac{z}{r}\phi,$$

$$H_x = 0, \quad H_y = -\beta\frac{v}{c}\frac{z}{r}\phi, \quad H_z = \beta\frac{v}{c}\frac{y}{r}\phi.$$

The electromagnetic energy and momentum are given by E and G respectively where

$$E = \frac{1}{2}\iiint(E_x^2 + E_y^2 + E_z^2 + H_x^2 + H_y^2 + H_z^2)dXdYdZ$$

$$G_x = \frac{1}{c}\iiint(E_yH_z - E_zH_y)dXdYdZ$$

$$G_y = \frac{1}{c}\iiint(E_zH_x - E_xH_z)dXdYdZ$$

$$G_z = \frac{1}{c}\iiint(E_xH_y - E_yH_x)dXdYdZ.$$

The integration is in each case through

the whole XYZ space *at a constant* T. The integrands are more simply expressed in terms of x, y, z, and we accordingly transform the integrals so that they are taken through the whole x, y, z space. Since $T = \beta(t + vx/c^2)$, the condition T = const. implies that t is not constant, so that the integrand in x, y, z must be taken for different values of t. In the particular case under consideration, however, the integrand is constant for all values of t, and no complication arises. Since $x = \beta(X + vT)$, $y = Y$, $z = Z$, we have for T = const.,

$$dXdYdZ = \beta^{-1}\, dxdydz.$$

We have

$$E = \frac{1}{2\beta}\iiint\{x^2 + \beta^2(y^2 + z^2) + \beta^2\frac{v^2}{c^2}(y^2 + z^2)\}\frac{\phi^2}{r^2}\,dxdydz.$$

Transforming to polar co-ordinates defined by

$$x = r\cos\theta,\ y = r\sin\theta\cos\phi,\ z = r\sin\theta\sin\phi,$$

we have

$$E = \frac{1}{2\beta}\int_o^{2\pi} d\phi \int_o^{\pi} d\theta \int_o^{\infty} \{\cos^2\theta +$$

$$\beta^2\left(1 + \frac{v^2}{c^2}\right)\sin^2\theta\} r^2\phi^2 \sin\theta dr$$

$$= \frac{\pi}{\beta}\left(\frac{2}{3} + 2 \cdot \frac{2}{3}\left(1 + \frac{v^2}{c^2}\right)\beta^2\right)\int_o^{\infty} r^2\phi^2 dr$$

$$= \frac{2}{3}\pi\beta\left(3 + \frac{v^2}{c^2}\right)\int_o^{\infty} r^2\phi^2 dr.$$

If W is the electrostatic energy of the electron when at rest, this may be calculated in the ordinary way or it may be obtained by putting $v = o$ in the above. Then

$$W = 2\pi \int_o^{\infty} r^2\phi^2 dr$$

or

$$E = \frac{1}{3}\beta\left(3 + \frac{v^2}{c^2}\right)W.$$

We may note that, neglecting fourth and higher powers of v/c, this gives

$$E = W + \frac{5}{6}\frac{W}{c^2}v^2.$$

If we identify the second term with the kinetic energy of the electron, and write it

$\frac{1}{2}mv^2$ where m is the electromagnetic mass of the electron, we have

$$m = \frac{5W}{3c^2}.$$

If we conceive the electron to be a sphere of radius a with a charge e spread uniformly over its surface, $W = \frac{1}{2}\frac{e^2}{a}$ and

$$m = \frac{5e^2}{6c^2a}.$$

On the other hand, if the charge is spread uniformly through the volume of the electron $W = \frac{3}{5}\frac{e^2}{a}$ and

$$m = \frac{e^2}{ac^2}.$$

Returning to the momentum, it is easily seen that $G_y = G_z = 0$, while

$$G_x = \frac{\beta v}{c^2}\iiint(y^2 + z^2)\frac{\phi^2}{r^2}\,dxdydz$$

$$= \frac{2\pi\beta v}{c^2} \cdot 2 \cdot \frac{2}{3}\int_0^\infty r^2\phi^2 dr$$

$$= \frac{4}{3}\frac{\beta v}{c^2}W.$$

If we make the improbable assumption that the field of an electron moving with variable velocity is at every instant the same as if the electron were moving with a constant velocity equal to its instantaneous velocity, then

$$\frac{d}{dt}(\mathrm{G}_x) = \frac{d}{dv}(\mathrm{G}_x)\frac{dv}{dt}$$

so that the longitudinal mass is given by

$$\frac{d}{dv}(\mathrm{G}_x) = \beta^3 \frac{4}{3}\frac{\mathrm{W}}{c^2}$$

as is easily seen to be true without approximation. The rest mass is obtained by dividing by β^3, and is accordingly

$$\frac{4}{3}\frac{\mathrm{W}}{c^2}.$$

V

THE EQUIVALENCE HYPOTHESIS

WE are now in a position to state the problem of relativity in its general form, and to indicate the kind of solution which Einstein has proposed.

The laws of physics need for their mathematical statement a set of axes in space, or a "frame of reference." If we take two such frames of reference, one moving relatively to the other, we should expect that the corresponding physical laws would be different. The classical view was that we should obtain the physical laws in their simplest form by choosing for our frame of reference a set of axes "at rest." The specification of this set of axes at rest has proved historically a matter of very great difficulty. All our physical observations are of relative motions, and cannot of themselves lead to the determination of absolute rest. One great historical effort to solve the dilemma was the development of the theory of the ether. Although not itself susceptible to physical observation, it might yet serve as a standard of absolute rest. Since the ether is the seat of all physical phenomena, the laws of physics might well assume a peculiar simplicity when they are stated with reference to a set of axes at rest in the ether. The proposed solution was briefly this: the absolute frame of reference is that for which Newton's laws of motion and Maxwell's laws of electromagnetism are accurately satisfied; the motion of any other frame of reference will be revealed by complication of these laws. The proposed solution failed because the Michelson and other experiments compelled us to assume that, if we take a moving frame of reference, the very motion of the frame gives rise to certain compensations which prevent us from detecting the motion.

The laws of physics are *not* more complicated for a "moving" frame than they are for a "fixed" frame; they are precisely the same. This failure drives us back to the original difficulty. We can find no meaning in physical experience for absolute motion, nor can we determine the frame of reference which is at rest.

It may be that experiment may yet discover some answer to the problem and some means of measuring our motion through space. At the same time, a great body of evidence suggests that it would be well to face the possibility of ultimate failure.

This is the standpoint of the theory of relativity. It assumes that, of the infinity of possible frames of reference, each moves relatively to the others, but none is "at rest" in any absolute or unique sense. We may select any frame of reference, but we must recognize that it is only one of an infinite number of equally eligible frames. We may by experiment determine the laws of physics for our selected frame, but they will be relative to that frame and, if we choose a different frame, the corresponding laws will be different. But we think of physical phenomena as pursuing their course independently of our measurement or description, and if this be so, there ought to be certain physical laws which are independent of the particular frame of reference which we may happen to have chosen. We thus arrive at the great problem of relativity: is it possible to express the laws of physics in a form which is independent of our choice of a frame of reference? Such laws of physics, if they exist, may well be called the absolute laws of physics.

The restricted theory of relativity has supplied a partial solution to the problem. If we confine ourselves to frames of reference which are moving relatively one to another with constant velocity in a straight line, we have seen, for example, that the equations of the electromagnetic field have precisely the same form for all such frames of reference. Thus in this restricted sense

Maxwell's equations express absolute laws of physics. This is clearly only a partial solution, which falls short of the full requirements of relativity. The restriction to frames of reference moving relatively one to another with uniform velocity was felt to be arbitrary, and many attempts were made to remove it. This was accomplished in a very complete manner by Einstein in his general theory of relativity. He showed that by taking gravitation into account the laws of physics may be expressed in the same form for all frames of reference.

In order to see how this was possible, we will examine briefly some of the outstanding features of gravitation. Newton interpreted gravitation as arising from the mutual attraction of bodies. Between any two bodies there is a force which is proportional to the product of their masses and inversely proportional to the square of their distance apart. Thus, according to Newton, gravitation was a mutual action between the attracting body and the body attracted.

Against this view we may note that the force is observable only through the acceleration which it produces in the attracted body, and this, being equal to the force divided by the mass, is independent of the body attracted. Just as Maxwell transferred the emphasis from attracting charges to the electromagnetic field, so Einstein directed attention to the gravitational field itself rather than to attracting bodies. This change of view-point brings to light a fundamental simplicity of gravitational fields which was somewhat obscured by the Newtonian presentation. A gravitational field impresses upon a body placed in it an acceleration which is quite independent of the body itself. Thus the uniform gravitational field which we experience in a limited region at the earth's surface means that *any* body free to move in it has a downward acceleration of approximately 32 ft./sec.². The gravitational field of the sun means that a planet at a given distance from the sun has an acceleration which depends on the sun and not on the planet. It was in this description of gravitation in terms of accelerations that Einstein found the way to the extension of relativity.

In the first place, it suggests a means by which all the appearance of a gravitational field may be produced artificially. Suppose that there is no gravitation but that an observer works in a room which is moving "upwards" with an acceleration g. All his observations inside the room will lead him to the conclusion that there is a gravitational field of the type familiar to us at the earth's surface. A body left free to move will in reality remain at rest or in uniform motion in a straight line. Suppose it is at rest. It will appear to the observer to fall downwards with an acceleration g which is the same for all bodies. If he projects a particle, it will appear to describe a parabola. A pendulum would execute oscillations in conformity with the usual formula. In short, by every test that the observer could make, he is at rest in a uniform gravitational field. The classical view draws a sharp distinction between an "artificial" gravitational field of this kind and a "true" gravitational field. Relativity denies the distinction because it cannot be tested by physical experiment. It denies that the observer and his room are moving in any absolute sense, but suggests rather that the observed facts may be interpreted in, among others, two ways—(1) the room is at rest and is occupied by a uniform gravitational field of intensity g; (2) there is no gravitational field, but the room is moving with an acceleration g. This liberty of interpretation is the essence of Einstein's "equivalence hypothesis." It does not imply that gravitation is merely an appearance arising from acceleration of our frame of reference, neither does it imply that for any given problem the two interpretations are equally simple or convenient. It merely insists that the two interpretations are equally true to the observable facts. It points the way to a complete solution of the problem of the choice of a frame of reference. We may select *any* frame of reference and observe the gravitational field ; we may with equal justification select any other frame of reference, but the gravitational field will then, in general, be different. All frames of reference are equally valid ; the only distinction between them is that for one the gravitational field may be simpler than for another. Thus, by bringing gravitation into account, Einstein was able to extend the theory of relativity to systems in relative acceleration.

It should be pointed out that the problem is not always so simple as in the case of the uniform gravitational field considered above. The gravitational field of the earth as a whole cannot be interpreted on the assumption that the radius of the earth is increasing at an accelerated rate of 32 ft./sec.^{2}. This merely means that there are some gravitational fields such that, of the infinity of possible frames of reference, there is no one for which the gravitation is everywhere nil.

We will return later to some discussion of the methods used in such cases.

The inclusion of gravitation in the new theory was the source of some of its greatest achievements, for gravitation had so far occupied a very isolated position in the scheme of physics. There appeared to be no interconnection between it and other physical phenomena. We will conclude this lecture by two examples of the way in which the liberty of interpretation permitted by the equivalence hypothesis enables us to establish such interconnections. The methods which we shall use are open to criticism at several points, but they serve to bring out the true nature of the equivalence hypothesis, and the results may be verified by more rigorous analysis.

Our first problem will be the effect of a gravitational field on the path of a ray of light. Suppose there are two sets of parallel axes $Oxyz$ and $O'x'y'z'$, and that O has an acceleration g relative to O′ in the positive direction of z, and that, if O′ is regarded as being at rest, there is no gravitation (see Fig. 3). Light is emitted from O and is received at a point P, on the axis of x, by a telescope inclined at an angle α to the axis of x. We will interpret this result firstly from the point of view of the "fixed"

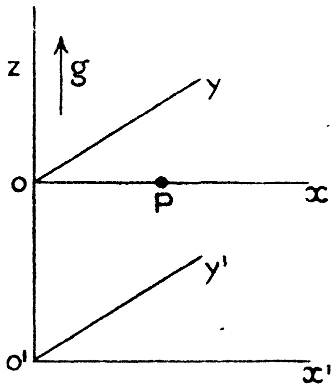

FIG. 3.

frame of reference $O'x'y'z$. There will then be no question of gravitation, but there will be an aberration effect due to the relative motion as between O at the instant of emission and P at the instant of reception. If c is the velocity of light and $OP = l$, this relative velocity is gl/c. The small angle of aberration will be this divided by the velocity of light, *i.e.*,

$$\alpha = gl/c^2.$$

We will now interpret the same phenomenon from the point of view of the frame of reference Oxyz. There will now be a field of gravitation of intensity g downwards. For simplicity, we will adopt a corpuscular theory of light, and admit the possibility that the corpuscles have weight, so that they have a downward acceleration G. Following the ordinary theory of projectiles, a corpuscle, projected from O and arriving at P with an angle of descent α, has its z-component of velocity reversed in time l/c approximately. Hence $l\text{G}/c = 2c \sin \alpha$, or approximately

$$\alpha = \text{G}l/2c^2.$$

Comparing the two interpretations we see that G = $2g$. That is to say, the light corpuscles are subject to the influence of gravitation, and experience an acceleration which is twice that experienced by a material particle.

A material comet, which at a great distance from the sun is moving with a high velocity c along a line whose perpendicular distance from the sun is p, is deflected by the sun's attraction through an angle $2\gamma M/pc^2$, where M is the mass of the sun and γ is the constant of gravitation. Accordingly, if we are justified in applying to the sun's gravitational field the result which we have proved above for a uniform gravitational field, a light corpuscle coming from a distant star will be deflected through an angle

$$\frac{4\gamma M}{pc^2}.$$

If the ray of light just grazes the limb of the sun, so that p is the sun's radius, this formula gives an angle of 1·73″, which agrees with the results obtained by observation.

Our second problem will be the effect of a gravitational field on the observed frequency of the radiation emitted by a vibrating atom. In Fig. 4 let $Oxyz$ and $O'x'y'z$ be two sets of parallel axes, and let O have an acceleration g relative to O′ in the direction of z. Let two precisely similar atoms, S_1 and S_2, be fixed relatively to O,

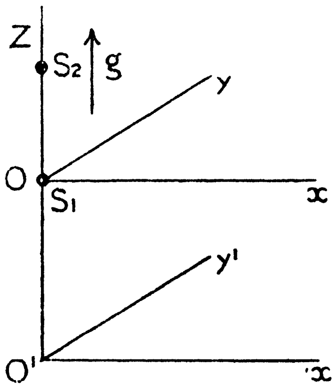

FIG. 4.

and situated at the points (o, o, o) and (o, o, z) respectively. The frequencies of their emitted radiation are observed at O′. For simplicity we shall assume that the relative velocity of O and O′ vanishes at the instant at which the figure is drawn. Firstly, we take the point of view that O′

is at rest, and that there is no gravitational field. Although the atoms are at rest relatively to O′ when the radiation is received, they had a relative velocity when the radiation was emitted, and this will give rise to a Doppler effect. If ν is the frequency of the emitted radiation, ν_1', ν_2' the respective observed frequencies at O′, and if $O'O = z_0$, the velocity of S_2 at the moment of emission was $(z + z_0')g/c$ towards O′ and hence

$$\nu_2' = \frac{\nu c}{c - (z + z_0)g/c}.$$

Since ν_1' is obtained from this by setting $z = 0$, we have

$$\frac{\nu_2'}{\nu_1'} = \frac{c - z_0 g/c}{c - (z + z_0)g/c}.$$

We will now calculate the same ratio from the point of view that O is at rest in a uniform gravitational field of intensity g in the negative direction of z. We will leave open the question as to whether the frequencies of the atoms are affected by the gravitational field and suppose that these are respectively ν_1 and ν_2. As O′ is at rest at the instant of reception, we have

$$\frac{\nu_2'}{\nu_1'} = \frac{\nu_2}{\nu_1}.$$

Comparing the values of the ratio calculated from the two points of view, we see that $\nu_1 \neq \nu_2$, and, in fact,

$$\frac{\nu_2}{\nu_1} = \frac{c - z_0 g/c}{c - (z + z_0)g/c}.$$

If ϕ is the gravitational potential with the arbitrary constant adjusted so that ϕ vanishes at O′, we have $\phi_1 = - z_0 g$, $\phi_2 = - (z + z_0)g$, and

$$\frac{\nu_2}{\nu_1} = \frac{1 + \phi_1/c^2}{1 + \phi_2/c^2},$$

or

$$\nu(1 + \phi/c^2) = \text{const.}$$

Accordingly the frequency of the radiation emitted by an atom in a place of high gravitational potential is less than that emitted by a similar atom in a place of low gravitational potential. Thus the lines of the solar spectrum should be displaced slightly towards the red as compared with the corresponding lines of atoms vibrating in the comparatively low gravitational potential of a terrestrial laboratory.

As we have already remarked, the solutions which we have given to these two problems are very open to criticism, but they serve to show the essence of Einstein's method, which is to describe one and the same physical phenomenon from the point of view of two frames of reference.

VI

THE FOUR-DIMENSIONAL CONTINUUM

OUR last lecture was devoted to a discussion of the powerful method by which Einstein brought gravitation into relation with other physical phenomena. But this was only a stepping-stone to the accomplishment of his main purpose, which was to supply a complete answer to the difficulty of the choice of a frame of reference by formulating the laws of physics in such a way that they are true for all possible frames. Before we can follow him further in this direction, we must make the acquaintance of some of the mathematical methods which he employed.

One of the most important of these is related to the idea of space-time. It is commonly stated that the theory of relativity assumes that space and time as we ordinarily understand them are not essentially distinct, but that they are merely special aspects of a more fundamental four-dimensional space-time. It has always seemed to me that this is a strong suspicion which might occur to one after a deep study of the theory rather than a dogma which must be accepted at the beginning. It is sufficient at this stage if we accept the four-dimensional continuum as a convenient mathematical representation. It is not necessary to assume that pressure and volume are fundamentally of the same nature before we can plot pressure against volume, or, as we may say, draw the isothermals of a gas in a two-dimensional pressure-volume space. In the same way we can represent an event occurring at a given place (x, y, z) at a given time (t) by a point (x, y, z, t) in a four-dimensional space.

It will be easier to grasp the simplicity of this representation if, for the moment, we reduce space to two dimensions and represent the motion of a point in the plane x, y by means of the three-dimensional space-time x, y, t. The career of a particle, moving with velocity u, v in a straight line and passing through the origin at time $t = 0$, is represented by the straight line

$$x/u = y/v = t.$$

A particle describing the circle $x^2 + y^2 = a^2$ with constant angular velocity ω is represented by the helix

$$x = a \cos \omega t, \ y = a \sin \omega t.$$

Whatever the motion of a particle in the plane x, y, its complete history, past, present, and future, will be represented by a "world-line" in the space x, y, t. The representation of the motion of a particle in three-dimensional space x, y, z requires a four-dimensional space-time x, y, z, t. The fact that we cannot represent such a space in three dimensions—that we cannot, for example, draw four lines each perpendicular to the other three—is of little more account than the familiar difficulty of representing three-dimensional figures on two-dimensional paper.

It being granted that the representation is possible, our next step is to show that it is convenient in that it supplies a simple interpretation of known results. For this purpose we will employ $\tau = ict$ for the fourth co-ordinate, where c is a constant ultimately identified with the velocity of light. The introduction of the imaginary i at this stage is by no means necessary, and is not resorted to in modern presentations of the theory. In our space x, y, z, τ, suppose the axes of x and τ are rotated through an angle θ, the other axes remaining as before. The formulæ of transformation for two axes about their origin are

$$x' = x \cos\theta + \tau \sin\theta, \; \tau' = -x \sin\theta + \tau \cos\theta.$$

Now suppose $\cos\theta = \beta$, $\sin\theta = iv\beta/c$, where $\beta^2 = 1/(1 - v^2/c^2)$; θ is then a pure imaginary. The formulæ become

$$x' = \beta(x + iv\tau/c), \; \tau' = \beta(\tau - ivx/c),$$

or, expressed in terms of t,

$$x' = \beta(x - vt), t' = \beta(t - vx/c^2).$$

Thus the Lorentz transformation admits of a simple, if somewhat strange, interpretation. It corresponds in the four-dimensional space x, y, z, τ to a mere transformation to new axes, obtained by rotating the axes of x, τ in their own plane through an imaginary angle.

Let us consider from this point of view the rather complicated formulæ for the transformation of velocities which we obtained in Lecture II. We have

$$u_x = \frac{dx}{dt} = ic\frac{dx}{d\tau},$$

so that if χ is the angle between the axis of x and the projection of the tangent to the world-line on the plane of $x\tau$, $u_x = ic \cot \chi$. Now let the axes of x and τ be turned through the angle θ, as described above If $\chi' = \chi - \theta$, the new component of velocity is

$$u_x' = ic \cot \chi' = \frac{ic(1 + \cot \chi \cot \theta)}{\cot \theta - \cot \chi}.$$

Substituting for χ and θ, this gives the result already established,

$$u_x' = \frac{u_x - v}{1 - vu_x/c^2}.$$

We are now in a position to see a reason for the complexity of this result. The components of velocity serve to define the direction of the tangent to the world-line. For this purpose we have used $u_x = ic \cot \chi$ and similar expressions for u_y and u_z. A more symmetrical procedure would be to employ the "direction-cosines" $\frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds}, \frac{d\tau}{ds}$, where ds is the length of an element of arc of the world-line and

$$ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2.$$

We accordingly take

$$ic\frac{dx}{ds}, \quad ic\frac{dy}{ds}, \quad ic\frac{dz}{ds}, \quad ic\frac{d\tau}{ds}$$

as the components of a four-dimensional velocity vector (the factor ic is inserted so that for slow motion these approximate to

$u_x, u_y, u_z, 1$). These suggest a four-dimensional acceleration vector with components

$$-c^2\frac{d^2x}{ds^2},\quad -c^2\frac{d^2y}{ds^2},\quad -c^2\frac{d^2z}{ds^2},\quad -c^2\frac{d^2\tau}{ds^2}.$$

The velocity and acceleration vectors are then simple aspects of the geometry of the world-line of the particle. Without implying too much by the names, we may say, on the analogy of the three-dimensional geometry of curves, that the velocity defines the direction of the tangent to the world-line, while the acceleration defines the magnitude and direction of its curvature.

Consider the equations

$$-mc^2\frac{d^2x}{ds^2} = k_1,$$

$$-mc^2\frac{d^2y}{ds^2} = k_2,$$

$$-mc^2\frac{d^2z}{ds^2} = k_3,$$

$$-mc^2\frac{d^2\tau}{ds^2} = k_4,$$

where m is the constant mass of the particle and k is a vector whose first three components are the components of force, while k_4 is at present undefined. If we can show that these equations express the law of motion of the particle, we may say that the four-dimensional view has restored to Newton's laws the simplicity which they seemed to lose under the Lorentz theory. To do this we must translate the above equations back into the ordinary three-dimensional notation. If $u_x = dx/dt$, . . ., and if $u_x^2 + u_y^2 + u_z^2 = v^2$, so that v is the resultant velocity in the ordinary sense, we have $ds^2 = (v^2 - c^2)dt^2$, or

$$\frac{d}{ds} = \frac{i\beta}{c}\frac{d}{dt}.$$

The first of the above equations then becomes

$$\frac{d}{dt}(\beta m u_x) = \beta^{-1}k_1.$$

This, together with the second and third equations, expresses the law of momentum with the variable mass βm.

Multiply the equations by dx/ds, . . ., and add. The left-hand side becomes the differential of a constant, and hence

$$k_1 \frac{dx}{ds} + k_2 \frac{dy}{ds} + k_3 \frac{dz}{ds} + k_4 \frac{d\tau}{ds} = 0,$$

or

$$k_4 = \frac{i}{c}(k_1 u_x + k_2 u_y + k_3 u_z).$$

The fourth equation then gives

$$\frac{d}{dt}(\beta mc^2) = \beta^{-1}(k_1 u_x + k_2 u_y + k_3 u_z).$$

This expresses the law of energy since $mc^2\beta$, or $mc^2(\beta - 1)$, is the form already found for the kinetic energy.

We cannot now follow the expression of the whole of the restricted theory of relativity in terms of four-dimensional space-time, but the above discussion of the theory of variable mass may serve to indicate the simplification which follows the adoption of this point of view. This was Minkowski's great contribution to the development of the theory of Relativity.*

* For a translation of Minkowski's classical memoir "Space and Time," *see* "The Principle of Relativity," pp. 75-96 (Methuen, 1923).

The presentation of the Lorentz transformation as the effect of a change of rectangular axes in four-dimensional space-time points the way to the formulation of the absolute laws of physics. We are familiar with the simplicity which is introduced into physics by the introduction of the idea of a vector. Let us return again to the law of motion. We may express this as follows: There is a certain vector, the mass acceleration, and there is another vector, the force, and these two vectors are equal in magnitude and direction. We usually express this law in a somewhat round-about way by resolving each vector along three directions and asserting the equality of the respective components. The equations thus obtained will depend upon our choice of axes, but the law in its vector form is independent of this choice. Thus it is in the development of the vector point of view that we shall seek the emancipation of the laws of physics. After what we have seen, we shall be prepared to find that the appropriate vectors are four-dimensional vectors in space-time rather than the ordinary three-dimensional vectors.

Our next step is to obtain an appropriate generalization of the idea of a vector to four dimensions. In three dimensions we define a vector as a quantity having magnitude and direction, and such that two vectors add by the parallelogram law. The fact that it will often be difficult to form images of the geometrical relations in four dimensions suggests that it will be convenient to express this definition in analytical rather than geometrical form. Suppose that there are two sets of rectangular axes (x_1, x_2, x_3) and (x_1', x_2', x_3') having a common origin and that the direction cosines of one set with respect to the other are such that the formulæ of transformation are

$$\begin{aligned} x_1' &= l_1 x_1 + m_1 x_2 + n_1 x_3 \\ x_2' &= l_2 x_1 + m_2 x_2 + n_2 x_3 \\ x_3' &= l_3 x_1 + m_3 x_2 + n_3 x_3 \end{aligned}$$

or, what is the same thing,

$$
\begin{aligned}
x_1 &= l_1 x_1' + l_2 x_2' + l_3 x_3' \\
x_2 &= m_1 x_1' + m_2 x_2' + m_3 x_3' \\
x_3 &= n_1 x_1' + n_2 x_2' + n_3 x_3'
\end{aligned}
$$

A vector which has components A_1, A_2, A_3 along the axes of x_1, x_2, x_3 will have components along x_1', x_2', x_3' equal to A_1', A_2', A_3', where

$$A_1' = l_1 A_1 + m_1 A_2 + n_1 A_3, \text{ etc.}$$

This may be expressed in either of the two forms—

$$A_1' = \frac{\partial x_1'}{\partial x_1} A_1 + \frac{\partial x_1'}{\partial x_2} A_2 + \frac{\partial x_1'}{\partial x_3} A_3,$$

or

$$A_1' = \frac{\partial x_1}{\partial x_1'} A_1 + \frac{\partial x_2}{\partial x_1'} A_2 + \frac{\partial x_3}{\partial x_1'} A_3.$$

Each of these is typical of three equations which may be written in condensed form—

$$\left.\begin{aligned}
A_\mu' &= \sum_{\nu=1}^{3} \frac{\partial x_\mu'}{\partial x_\nu} A_\nu \\
A_\mu' &= \sum_{\nu=1}^{3} \frac{\partial x_\nu}{\partial x_\mu'} A_\nu
\end{aligned}\right\} \quad (\mu = 1, 2, 3).$$

A three-dimensional vector may be defined as a quantity having three components A_1, A_2, A_3, which transform from one set of co-ordinates to another by either of the above formulæ of transformation. It can be shown that this analytical definition is equivalent to the more usual geometrical definition.

The extension to four dimensions is now obvious. Before stating it, we will note that the equivalence of the above two forms of the transformation law is a special property of transformations with rectangular axes. For other transformations it is necessary to distinguish between the vectors which obey the one or the other. This is done by the position of the index indicating the component. We define as follows :—

A Co-variant Vector is a quantity having four components A_μ ($\mu = 1, 2, 3, 4$), which transforms from one set of co-ordinates to another by the law

$$A_\mu' = \sum_{\nu=1}^{4} \frac{\partial x_\nu}{\partial x_\mu'} A_\nu.$$

A Contravariant Vector has four components $A^\mu(\mu = 1, 2, 3, 4)$, which transform by the law

$$A'^\mu = \sum_{\nu=1}^{4} \frac{\partial x_\mu'}{\partial x_\nu} A^\nu.$$

The distinction between the two types of vector does not arise if we restrict ourselves to changes of orthogonal axes.

Consider two of the vectors which we introduced earlier in the lecture. The velocity vector has components which we may now write as proportional to $dx_\nu/ds = A^\nu$. Then, taking the contravariant law,

$$A^{\mu'} = \sum_{\nu=1}^{4} \frac{\partial x_\mu'}{\partial x_\nu} \frac{dx_\nu}{ds} = \frac{dx_\mu}{ds}.$$

Hence the velocity is a contravariant vector.

On the other hand, the acceleration vector

$$B^{\mu'} \equiv \frac{d^2 x_\mu'}{ds^2} = \frac{d}{ds}\left\{\sum_{\nu=1}^{4} \frac{\partial x_\mu'}{\partial x_\nu} \frac{dx_\nu}{ds}\right\}$$

$$= \sum_{\nu=1}^{4} \frac{\partial x_\mu'}{\partial x_\nu} B^\nu + \sum_{\nu=1}^{4} \sum_{\rho=1}^{4} \frac{\partial^2 x_\mu'}{\partial x_\nu \partial x_\rho} \frac{dx_\nu}{ds} \frac{dx_\rho}{ds},$$

is not a vector in the full sense. However, for any transformation in which the new co-ordinates are linear functions of the old, the second differentials in the double sum vanish and the acceleration behaves as a contravariant vector. If, in addition, the transformation is orthogonal (like the Lorentz transformation), $\delta x_\mu'/\delta x_\nu = dx_\nu/dx_\mu'$ and the acceleration behaves as either a contravariant or a covariant vector.

In three dimensions we sometimes meet quantities with properties similar to those of a vector, but having more than three components. For example, if in an elastic solid $\widehat{xy}$ denotes the y-component of force across a unit area normal to the x-axis, the stress is completely specified by nine (3^2) components, $\widehat{xx}$, $\widehat{xy}$, . . . , some of which are equal. A typical equation for their transformation for rectangular axes is

$$\widehat{x'y'} = l_1 l_2 \widehat{xx} + m_1 m_2 \widehat{yy} + n_1 n_2 \widehat{zz} + (m_1 n_2 + m_2 n_1)\widehat{yz} + (n_1 l_2 + n_2 l_1)\widehat{zx} + (l_1 m_2 + l_2 m_1)\widehat{xy}.$$

If we write $\widehat{xx} = S_{11}$, $\widehat{xy} = S_{12}$, . . . , this becomes

$$S'_{\mu\nu} = \sum_{\alpha=1}^{3}\sum_{\beta=1}^{3} \frac{\partial x_\alpha}{\partial x_\mu'} \frac{\partial x_\beta}{\partial x_\nu'} S_{\alpha\beta}.$$

This suggests a definition for a " tensor " of the second rank as a quantity with $4^2 = 16$ components, which transforms according to a law corresponding to its character. A tensor of the second rank may be covariant, contravariant, or mixed. The law of transformation for a mixed tensor is

$$A_\nu^{\mu\prime} = \sum_{\alpha=1}^{4}\sum_{\beta=1}^{4} \frac{\partial x_\mu'}{\partial x_\alpha} \frac{\partial x_\beta}{\partial x_\nu'} A_\beta^\alpha.$$

From this it is easy to frame the definitions and laws of transformation of a covariant $(A_{\mu\nu})$ or a contravariant $(A^{\mu\nu})$ tensor of the second rank, or of a tensor of any higher rank or character (e.g., $A_{\rho\sigma\tau}^{\mu\nu}$). Vectors take their place as tensors of the first rank, and we may complete the scheme by defining a tensor of zero rank as a quantity with a single component which is unchanged by any change of co-ordinates—an invariant.

It is easy to prove from the definitions that the components of two tensors of the same rank and character may be added (or subtracted) to form the components of a new tensor. Rules can be laid down for the multiplication of tensors, and a consistent algebra of tensors can be built up.

The application of the theory of tensors to the problem of relativity arises in this way. Suppose it is possible to express a physical law by the equality of two tensors of the same rank and character, e.g. $mA^{\mu} = F^{\mu}$, where A^{μ} is the acceleration vector and F^{μ} the force vector. By transferring both terms to the same side of the equation, and using the rules for addition of tensors, the law may be expressed by saying that a certain tensor vanishes, that is to say, that each of its components vanishes. If we now take a new system of co-ordinates, each of the new components of the tensor is the sum of multiples of the old components, and therefore vanishes. If the components of a tensor vanish in one system of co-ordinates, they will vanish in all systems.

The way to make the laws of physics independent of our choice of co-ordinates is to express these laws as relations between tensors.

VII

THE GENERAL THEORY

IN the last lecture we saw that, from the point of view of four-dimensional space-time, the Lorentz transformation corresponds to a change of rectangular axes in which two of the axes are rotated through a certain angle in their own plane, while the other two are unaltered. From the ordinary three-dimensional point of view it corresponds to a change from one frame of reference to another, which moves relatively to the first with uniform velocity in the direction of one of the axes. The problem of relativity demands a much more general change for the frame of reference, namely, from one to another moving relatively to it in *any* manner. It might be thought that such a general change of the frame of reference might correspond to a more general change of axes in space-time. However, so long as we confine ourselves to rectangular Cartesian axes in space-time, this is not the case. The most general change of rectangular axes about the same origin corresponds to frames of reference which are moving relatively the one to the other with a velocity which is uniform but not necessarily in the direction of any of the axes ; a change to parallel axes through a new origin merely alters the origin in space and the zero instant in time. The most general change of rectangular axes in space-time thus gives no more than the Lorentz transformation.

If we look back over the last lecture we see how little depends upon the axes being rectilinear and rectangular, and Einstein's next step was to pass at once to a much more general conception of co-ordinates in space-time. Just as we frequently employ curvilinear co-ordinates for special purposes in three-dimensional space, so we employ curvilinear co-ordinates in space-time. In place of x, y, z, t we may take any four independent functions of these, x_1, x_2, x_3, x_4, and these will serve to specify the point of space-time corresponding to a physical event. These relations may be expressed

$$\left.\begin{array}{l} x = f_1(x_1, x_2, x_3, x_4) \\ y = f_2(x_1, x_2, x_3, x_4) \\ z = f_3(x_1, x_2, x_3, x_4) \\ t = f_4(x_1, x_2, x_3, x_4) \end{array}\right\} \quad . \quad . \quad (1)$$

If f_1, f_2, f_3, f_4 are arbitrary functions, we have the most general system of co-ordinates in space-time. The theory of tensors developed in the last lecture is applicable without modification. Physical laws, expressed in tensor form, are equally valid whether the co-ordinates are x, y, z, t or x_1, x_2, x_3, x_4.

In the development of the theory great importance attaches to the form for ds^2, the square of the length of an element of arc in space-time. Expressing it in terms of t instead of τ, and making an unimportant change of sign, it is

$$ds^2 = -dx^2 - dy^2 - dz^2 + c^2dt^2 \quad . \quad (2)$$

From (1) we obtain equations of the type

$$dx = \frac{\partial f_1}{\partial x_1} dx_1 + \frac{\partial f_1}{\partial x_2} dx_2 + \frac{\partial f_1}{\partial x_3} dx_3 + \frac{\partial f_1}{\partial x_4} dx_4,$$

and these may be used to transform (2) to the new variables. We obtain

$$ds^2 = \sum_{\mu=1}^{4} \sum_{\nu=1}^{4} g_{\mu\nu} dx_\mu dx_\nu \quad . \quad . \quad (3)$$

There is no loss of generality in assuming that $g_{\mu\nu} = g_{\nu\mu}$, and these coefficients are then definite functions of the differential coefficients of f_1, f_2, f_3, f_4, and are therefore functions of x_1, x_2, x_3, x_4.

It has been known since the time of Riemann that the form for ds^2 can be used as the basis for the geometry of the space to which it belongs. This method stands in sharp contrast to the more familiar methods of Euclidean and projective geometry. In these "finite" geometries we start with definitions of figures such as lines, planes, spheres, etc., and there is usually the underlying assumption that the space is homogeneous ; that its properties are everywhere exactly the same. Differential geometry, on the other hand, starts with the relations between infinitely near points as expressed by the form for ds^2, and, provided that suitable assumptions are made, arrives at conclusions consistent with those of the finite geometries. But differential geometry has this advantage, that it is much more readily applicable to the study of heterogeneous space whose properties are different from point to point. There is nothing strange in the idea of heterogeneous space. The surface of an ellipsoid is a two-dimensional heterogeneous space, and the surface of the earth is more so. The form (2) corresponds to a space which is homogeneous and, moreover, with respect to its three dimensions x, y, z, isotropic. In (3) the coefficients are functions of the coordinates and the space is heterogeneous.

The space-time of Einstein is heterogeneous. We have pointed out that gravitation is an effect by which any body placed in a given position acquires an acceleration which depends upon the position and not upon the accelerated body. Thus, in a sense, we may say that gravitation is a property of space-time and that the latter is therefore obviously heterogeneous. Some would prefer to say that the space-time considered by Einstein is not so much space-time as space-time cum gravitation.

It is easily proved, by making a further change of co-ordinates in (3), that the $g_{\mu\nu}$ are the components of a covariant tensor of the second rank. This tensor defines the gravitational field, and its components are sometimes spoken of as the gravitational potentials.

The homogeneous and isotropic form (2) can correspond only to a space in which there is no gravitation, and, since (3) has been obtained from (2) by a change of co-ordinates, the gravitation expressed by (3) may be said to be fictitious, and to be an appearance arising from our choice of co-ordinates. Thus, when we were considering a uniform gravitational field, we were able to choose a new frame of reference in such a way that the gravitation disappeared. This is not always possible. A substitution will always bring the form (2) to the form (3), but it does not follow that, if we start from (3) with given functions of the co-ordinates for $g_{\mu\nu}$, we can find a substitution which will bring us back to (2). This may be expressed by saying that there are some gravitational fields which cannot be "transformed away" by a change in the frame of reference. If we have a given gravitational field expressed by (3), it is clearly a question of fact, independent of the co-ordinates employed, as to whether there is a frame of reference for which ds^2 is given by (2). Accordingly this condition ought to be expressible by the vanishing of a tensor. This is the case, and the condition is $B^{\lambda}_{\mu\nu\rho} = 0$, where $B^{\lambda}_{\mu\nu\rho}$ is a mixed tensor of the fourth rank, called the Riemann-Christoffel tensor. It has $4^4 = 256$ components, but fortunately only twenty of these are independent and these are known functions of the $g_{\mu\nu}$ and their differential coefficients.

The general method of applying the equivalence hypothesis to this theory is as follows: Suppose we know some physical law for the case when there is no gravitational field. We express it in tensor (or invariant) form in the variables x, y, z, t applicable to the space (2). We now make the change of co-ordinates which brings (2) to (3), and we shall expect that the $g_{\mu\nu}$, expressing the resultant "fictitious" gravitational field, will appear in our tensor equations. Einstein assumes that the equations thus obtained will be valid even when the gravitational field is not "fictitious"; that the $g_{\mu\nu}$ enter into the expression of physical laws in precisely the same way whether they represent a gravitational field which can be "transformed away" or whether they represent one which cannot.

For example, the law of motion for a freely moving particle when there is no gravitational field is that it moves with uniform velocity in a straight line. We have seen that this motion corresponds to a straight world-line in space-time. Adopting the definition of a straight line as the shortest distance between two points, we may express the law by saying that $\int_A^B ds$ is a minimum where the integral is taken along the world-line between any two points A and B. Again, this may be written

$$\delta \int_A^B ds = 0, \quad . \quad . \quad . \quad (4)$$

where the left-hand side means the change in the value of the integral when it is taken along a curve joining A and B, but differing slightly from the world-line. Since ds is an invariant, the condition expressed in this form is independent of the co-ordinates employed. Accordingly Einstein assumes (4) to be the law of motion for all systems of co-ordinates, and for any gravitational field. The integral form (4) may be replaced by equivalent differential equations. If this is done it will be found, as we should expect, that these equations express the vanishing of a tensor. The carrying through of this work requires some slight further development of the theory of tensors, but we will show that (4) provides a practicable method of determining the motion of a particle by applying it to the form of equation (3), which Einstein found for the gravitational field of the sun.

This is most conveniently expressed in spherical polar co-ordinates r, θ, ϕ, and then

$$ds^2 = -\gamma^{-1}dr^2 - r^2d\theta^2 - r^2 \sin^2 \theta d\phi^2 + \gamma c^2 dt^2 \quad . \quad (5)$$

where

$$\gamma = 1 - \frac{2\kappa M}{c^2 r},$$

and M is the mass of the sun and κ the constant of gravitation.

If the particle (by which, of course, we mean a planet) is moving in the equatorial plane $\theta = \frac{\pi}{2}$, $d\theta = o$ and (4) may be written

$$\delta\int_A^B \sqrt{\left\{\gamma c^2\left(\frac{dt}{ds}\right)^2 - r^2\left(\frac{d\phi}{ds}\right)^2 - \frac{1}{\gamma}\left(\frac{dr}{ds}\right)^2\right\}}ds = 0. \quad (6)$$

It will simplify our formulæ if we consider the special cases of this equation in which ϕ and t separately have values slightly different from those appropriate to the actual path. First let ϕ be subject to a small variation $\delta\phi$, then $\delta\left(\frac{d\phi}{ds}\right) = \frac{d}{ds}(\delta\phi)$, and (6) gives

$$\int_A^B \frac{r^2\frac{d\phi}{ds}\frac{d}{ds}(\delta\phi)ds}{\sqrt{\left\{\gamma c^2\left(\frac{dt}{ds}\right)^2 - r^2\left(\frac{d\phi}{ds}\right)^2 - \frac{1}{\gamma}\left(\frac{dr}{ds}\right)^2\right\}}} = 0.$$

The denominator is equal to unity by (5), and hence

$$\int_A^B r^2\frac{d\phi}{ds}\frac{d}{ds}(\delta\phi)ds = 0.$$

Integrating by parts,

$$\left| r^2\frac{d\phi}{ds}\delta\phi\right|_A^B - \int_A^B\frac{d}{ds}\left(r^2\frac{d\phi}{ds}\right)\delta\phi ds = 0.$$

The variations of the path being subject to the condition that the path passes through A and B, we have $\delta\phi = 0$ at A and at B. Hence the first term vanishes. Apart from these terminal values, $\delta\phi$ is arbitrary, and hence the integral will vanish only if

$$\frac{d}{ds}\left(r^2 \frac{d\phi}{ds}\right) = 0.$$

or

$$r^2 \frac{d\phi}{ds} = \frac{h}{c}, \qquad . \qquad . \qquad (7)$$

where h is a constant.

Allowing t to take a small variation, and following precisely the same method, we find

$$\frac{d}{ds}\left(\gamma \frac{dt}{ds}\right) = 0,$$

or

$$\gamma \frac{dt}{ds} = C, \qquad . \qquad . \qquad . \qquad (8)$$

where C is a constant.

A third equation can be obtained by allowing r to take a small variation; this equation is more complicated, and we can proceed without it. For the case, $\theta = \frac{1}{2}\pi$, (5) is equivalent to

$$\left(\frac{dr}{ds}\right)^2 + \gamma r^2 \left(\frac{d\phi}{ds}\right)^2 = \gamma^2 c^2 \left(\frac{dt}{ds}\right)^2 - \gamma \quad . \quad (9)$$

Using (7) and (8) to eliminate s and t, writing $1/r = u$, and remembering the value of γ, this equation becomes

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{c^4 C^2}{h^2} - \frac{c^2}{h^2} + \frac{2\kappa M}{h^2} u + \frac{2\kappa M}{c^2} u^3. \quad (10)$$

Differentiating with respect to ϕ, and dividing by $2du/d\phi$, we have

$$\frac{d^2 u}{d\phi^2} + u = \frac{\kappa M}{h^2} + \frac{3\kappa M}{c^2} u^2 \quad . \quad (11)$$

The second term on the right-hand side of this equation is, in practical cases, very small compared with the first. For a circular orbit the ratio is three times the square of the ratio of the velocity of the planet to that of light. If, for the moment, we neglect this small term, the equation is identical with the well-known differential equation for central orbits under a central force $\kappa M/r^2$, with h for Kepler's areal constant. Thus an approximation to Newton's law of the inverse square appears as a consequence of Einstein's method. It is important to notice that we did not postulate any "force of gravitational attraction" in order to obtain this result. The particle was supposed to be unacted upon by any force, and to move "with uniform velocity in a straight line," but the idea of a straight line was modified to allow for the heterogeneity of space-time which marks the existence of a gravitational field. This non-appearance of gravitational force is characteristic of Einstein's theory. For the purpose of comparison with the older theory, however, we may interpret Einstein's result in terms of gravitational force. We will consider a simple case in which the particle moves in a straight line through the origin, so that $d\phi/ds = 0$. Equation (9) then becomes

$$\left(\frac{dr}{ds}\right)^2 = c^2C^2 - 1 + \frac{2\kappa M}{c^2 r}.$$

On differentiation this becomes

$$\frac{d^2r}{ds^2} = -\frac{\kappa M}{c^2r^2}.$$

If the left-hand side is taken as the acceleration, we have an accurate inverse square law; but if we take the more usual d^2r/dt^2 as the acceleration and calculate its value by the aid of (8), it will be found to contain a term varying as the inverse square of the distance, and, in addition, terms varying as the inverse third and fourth powers.

The second term on the right-hand side of (11), though small, has important consequences. Its effect is that the orbit is not accurately a conic, but may be represented by, for example, an ellipse whose major axis rotates slowly.

Let u_1, u_2 be two of the values of u for which $du/d\phi$ vanishes; they are the reciprocals of the apsidal distances. The constants C and h can be expressed in terms of u_1, u_2, and (10) becomes

$$\left(\frac{du}{d\phi}\right)^2 = (u_1 - u)(u - u_2)$$
$$\{1 - 2\kappa M(u + u_1 + u_2)/c^2\}.$$

$2\kappa Mu/c^2$ being small, this gives approximately

$$\phi = \int \frac{\{1 + \kappa M(u + u_1 + u_2)/c^2\}du}{\sqrt{\{(u_1 - u_1)(u - u_2)\}}}.$$

If we write $u = u_1 \cos^2 \psi + u_2 \sin^2 \psi$, then as ψ goes from 0 to π, u starts from its apsidal value u_1, and returns to that value. The angle ϕ described between corresponding apsidal positions is thus

$$2\int_0^\pi \{1 + \frac{\kappa M}{c^2}(u_1 + u_2 + u_1 \cos^2 \psi + u_2 \sin^2 \psi)\}d\psi$$
$$= 2\pi + \frac{3\pi\kappa M}{c^2}(u_1 + u_2).$$

The excess of this value over four right-angles represents a rotation of the apse line which has occurred during one revolution. If a, e are respectively the major semi-axis and eccentricity of the orbit, $u_1 = 1/a(1 - e)$, $u_2 = 1/a(1 + e)$, and $u_1 + u_2 = 2/a(1 - e^2) = 2/l$, where l is the semi-latus rectum. The rotation of the apse line for one revolution is thus $6\pi\kappa M/c^2l$. This result has been verified by observations on the orbit of the planet Mercury, for which it amounts to about 43″ per century.

We must now bring our account of the theory of relativity to a close. We have endeavoured to trace the development of the great work of Fresnel, and to show how the difficulty of determining an absolute frame of reference dogged the footsteps of science, and to explain in some measure the way in which Einstein has supplied an answer to that difficulty. It is clear that the subject merits a more thorough and systematic treatment than we have been able to give in these lectures. There is a mistaken idea that this treatment can be accomplished only by the aid of very difficult mathematics. It is true that one must master the theory of tensors and gain a certain facility in their manipulation, but this is by no means so difficult as is commonly supposed. A very clear, if somewhat compressed, account of the mathematical theory will be found in Eddington's " Report to the Physical Society on the Relativity Theory of Gravitation " (Fleetway Press); a fuller and more systematic treatment is given by the same author in " The Mathematical Theory of Relativity " (Cambridge University Press); translations of the classical papers on the subject by Lorentz, Einstein, Minkowski, and Weyl are published under the title of " The Principle of Relativity " (Methuen). Every young English physicist should study at least the first of these books, and it is hoped that these lectures will help to smooth the way by a preliminary exploration of the ground.

INDEX

PRINTED IN GREAT BRITAIN AT
THE UNIVERSITY PRESS, ABERDEEN