MEANING
OF
RELATIVITY

ALBERT
EINSTEIN

METHUEN

ALBERT EINSTEIN

THE MEANING OF RELATIVITY

Sixth, Definitive Edition
including the
RELATIVISTIC THEORY OF
THE NON-SYMMETRIC FIELD

First Published 2 November 1922
Second Edition October 1937
Third Edition with an Appendix October 1946
Fourth Edition with further Appendix April 1950
Fifth Edition July 1951
Sixth Edition, revised, 1956

6.1

CATALOGUE NO 3551/U

PRINTED AND BOUND IN GREAT BRITAIN BY
BUTLER AND TANNER LTD, FROME AND LONDON

This book, originally published in 1922, consisted of the text of Dr. Einstein's Stafford Little Lectures, delivered in May 1921 at Princeton University. For the third edition, Dr. Einstein added an appendix discussing certain advances in the theory of relativity since 1921. To the fourth edition, Dr. Einstein added Appendix II on his Generalized Theory of Gravitation. In the fifth edition the proof in Appendix II was revised.

In the present (sixth) edition Appendix II has been rewritten. This edition and the Princeton University Press fifth edition, revised (1955), are identical.

The text of the first edition was translated by Edwin Plimpton Adams, the first appendix by Ernst G. Straus and the second appendix by Sonja Bargmann.

THE previous considerations concerning the configuration of rigid bodies have been founded, irrespective of the assumption as to the validity of the Euclidean geometry, upon the hypothesis that all directions in space, or all configurations of Cartesian systems of co-ordinates, are physically equivalent. We may express this as the 'principle of relativity with respect to direction', and it has been shown how equations (laws of nature) may be found, in accord with this principle, by the aid of the calculus of tensors. We now inquire whether there is a relativity with respect to the state of motion of the space of reference; in other words, whether there are spaces of reference in motion relatively to each other which are physically equivalent. From the standpoint of mechanics it appears that equivalent spaces of reference do exist. For experiments upon the earth tell us nothing of the fact that we are moving about the sun with a velocity of approximately 30 kilometres a second. On the other hand, this physical equivalence does not seem to hold for spaces of reference in arbitrary motion; for mechanical effects do not seem to be subject to the same laws in a jolting railway train as in one moving with uniform velocity; the rotation of the earth must be considered in writing down the equations of motion relatively to the earth. It appears, therefore, as if there were Cartesian systems of co-ordinates, the so-called inertial systems, with reference to which the laws of mechanics (more generally the laws of physics) are expressed in the simplest form. We may surmise the validity of the following proposition: if K is an inertial system, then every other system K' which moves uniformly and without rotation relatively to K, is also an inertial system; the laws of nature are in concordance for all inertial systems. This statement we shall call the 'principle of special relativity'. We shall draw certain conclusions from this principle of 'relativity of translation' just as we have already done for relativity of direction.

In order to be able to do this, we must first solve the following problem. If we are given the Cartesian co-ordinates, x_ν, and the time, t, of an event relatively to one inertial system, K, how can we calculate the co-ordinates, x'_ν, and the time, t', of the same event relatively to an inertial system K' which moves with uniform translation relatively to K? In the pre-relativity physics this problem was solved by making unconsciously two hypotheses:

1. Time is absolute; the time of an event, t', relatively to K' is the same as the time relatively to K. If instantaneous signals could be sent to a distance, and if one knew that the state of motion of a clock had no influence on its rate, then this assumption would be physically validated. For then clocks, similar to one another, and regulated alike, could be distributed over the systems K and K', at rest relatively to them, and their indications would be independent of the state of motion of the systems; the time of an event would then be given by the clock in its immediate neighbourhood.

2. Length is absolute; if an interval, at rest relatively to K, has a length s, then it has the same length s, relatively to a system K' which is in motion relatively to K.

If the axes of K and K' are parallel to each other, a simple calculation based on these two assumptions, gives the equations of transformation

$$\left.\begin{aligned} x'_\nu &= x_\nu - a_\nu - b_\nu t \\ t' &= t - b \end{aligned}\right\} \qquad (21)$$

This transformation is known as the 'Galilean Transformation'. Differentiating twice by the time, we get

$$\frac{d^2x'_\nu}{dt^2} = \frac{d^2x_\nu}{dt^2}$$

Further, it follows that for two simultaneous events,

$$x'^{(1)}_\nu - x'^{(2)}_\nu = x^{(1)}_\nu - x^{(2)}_\nu$$

The invariance of the distance between the two points results from squaring and adding. From this easily follows the co-variance of Newton's equations of motion with respect to the Galilean transformation (21). Hence it follows that classical mechanics is in accord with the principle of special relativity if the two hypotheses respecting scales and clocks are made.

But this attempt to found relativity of translation upon the Galilean transformation fails when applied to electromagnetic phenomena. The Maxwell-Lorentz electromagnetic equations are not co-variant with respect to the Galilean transformation. In particular, we note, by (21), that a ray of light which referred to K has a velocity c, has a different velocity referred to K', depending upon its direction. The space of reference of K is therefore distinguished, with respect to its physical properties, from all spaces of reference which are in motion relatively to it (quiescent ether). But all experiments have shown that electro-magnetic and optical phenomena, relatively to the earth as the body of reference, are not influenced by the translational velocity of the earth. The most important of these experiments are those of Michelson and Morley, which I shall assume are known. The validity of the principle of special relativity with respect to electromagnetic phenomena also can therefore hardly be doubted.

On the other hand, the Maxwell-Lorentz equations have proved their validity in the treatment of optical problems in moving bodies. No other theory has satisfactorily explained the facts of aberration, the propagation of light in moving bodies (Fizeau), and phenomena observed in double stars (De Sitter). The consequence of the Maxwell-Lorentz equations that in a vacuum light is propagated with the velocity c, at least with respect to a definite inertial system K, must therefore be regarded as proved. According to the principle of special relativity, we must also assume the truth of this principle for every other inertial system.

Before we draw any conclusions from these two principles we must first review the physical significance of the concepts 'time' and 'velocity'. It follows from what has gone before, that co-ordinates with respect to an inertial system are physically defined by means of measurements and constructions with the aid of rigid bodies. In order to measure time, we have supposed a clock, U, present somewhere, at rest relatively to K. But we cannot fix the time, by means of this clock, of an event whose distance from the clock is not negligible; for there are no 'instantaneous signals' that we can use in order to compare the time of the event with that of the clock. In order to complete the definition of time we may employ the principle of the constancy of the velocity of light in a vacuum. Let us suppose that we place similar clocks at points of the system K, at rest relatively to it, and regulated according to the following scheme. A ray of light is sent out from one of the clocks, U_m, at the instant when it indicates the time t_m, and travels through a vacuum a distance r_{mn}, to the clock U_n; at the instant when this ray meets the clock U_n the latter is set to indicate the time $t_n = t_m + \frac{r_{mn}}{c}$.* The principle

* Strictly speaking, it would be more correct to define simultaneity first, somewhat as follows: two events taking place at the points A and B of the system K are simultaneous if they appear at the same instant when osberved from the middle point, M, of the interval AB. Time is then defined as the ensemble of the indications of similar clocks, at rest relatively to K, which register the same simultaneously.

of the constancy of the velocity of light then states that this adjustment of the clocks will not lead to contradictions. With clocks so adjusted, we can assign the time to events which take place near any one of them. It is essential to note that this definition of time relates only to the inertial system K, since we have used a system of clocks at rest relatively to K. The assumption which was made in the pre-relativity physics of the absolute character of time (i.e. the independence of time of the choice of the inertial system) does not follow at all from this definition.

The theory of relativity is often criticized for giving, without justification, a central theoretical rôle to the propagation of light, in that it founds the concept of time upon the law of propagation of light. The situation, however, is somewhat as follows. In order to give physical significance to the concept of time, processes of some kind are required which enable relations to be established between different places. It is immaterial what kind of processes one chooses for such a definition of time. It is advantageous, however, for the theory, to choose only those processes concerning which we know something certain. This holds for the propagation of light *in vacuo* in a higher degree than for any other process which could be considered, thanks to the investigations of Maxwell and H. A. Lorentz.

From all of these considerations, space and time data have a physically real, and not a mere fictitious, significance; in particular this holds for all the relations in which co-ordinates and time enter, e.g. the relations (21). There is, therefore, sense in asking whether those equations are true or not, as well as in asking what the true equations of transformation are by which we pass from one inertial system K to another, K', moving relatively to it. It may be shown that this is uniquely settled by means of the principle of the constancy of the velocity of light and the principle of special relativity.

To this end we think of space and time physically defined with respect to two inertial systems, K and K', in the way that has been shown. Further, let a ray of light pass from one point P_1 to another point P_2 of K through a vacuum. If r is the measured distance between the two points, then the propagation of light must satisfy the equation

$$r = c . \Delta t$$

If we square this equation, and express r^2 by the differences of the co-ordinates, Δx_ν, in place of this equation we can write

$$\Sigma(\Delta x_\nu)^2 - c^2 \Delta t^2 = 0 \quad . \quad . \quad (22)$$

This equation formulates the principle of the constancy of the velocity of light relatively to K. It must hold whatever may be the motion of the source which emits the ray of light.

The same propagation of light may also be considered relatively to K', in which case also the principle of the constancy of the velocity of light must be satisfied. Therefore, with respect to K', we have the equation

$$\Sigma(\Delta x'_\nu)^2 - c^2 \Delta t'^2 = 0 \quad . \quad . \quad (22a)$$

Equations (22a) and (22) must be mutually consistent with each other with respect to the transformation which transforms from K to K'. A transformation which effects this we shall call a 'Lorentz transformation'.

Before considering these transformations in detail we shall make a few general remarks about space and time. In the pre-relativity physics space and time were separate entities. Specifications of time were independent of the choice of the space of reference. The Newtonian mechanics was relative with respect to the space of reference, so that, e.g. the statement that two non-simultaneous events happened at the same place had no objective meaning (that is, independent of the space of reference). But this relativity had no rôle in building up the theory. One spoke of points of space, as of instants of time, as if they were absolute realities. It was not observed that the true element of the space-time specification was the event specified by the four numbers x_1, x_2, x_3, t. The conception of something happening was always that of a four-dimensional continuum; but the recognition of this was obscured by the absolute character of the pre-relativity time. Upon giving up the hypothesis of the absolute character of time, particularly that of simultaneity, the four-dimensionality of the time-space concept was immediately recognized. It is neither the point in space, nor the instant in time, at which something happens that has physical reality, but only the event itself. There is no absolute (independent of the space of reference) relation in space, and no absolute relation in time between two events, but there is an absolute (independent of the space of reference) relation in space and time, as will appear in the sequel. The circumstance that there is no objective rational division of the four-dimensional continuum into a three-dimensional space and a one-dimensional time continuum indicates that the laws of nature will assume a form which is logically most satisfactory when expressed as laws in the four-dimensional space-time continuum. Upon this depends the great advance in method which the theory of relativity owes to Minkowski. Considered from this standpoint, we must regard x_1, x_2, x_3, t as the four co-ordinates of an event in the four-dimensional continuum. We have far less success in picturing to ourselves relations in this four-dimensional continuum than in the three-dimensional Euclidean continuum; but it must be emphasized that even in the Euclidean three-dimensional geometry its concepts and relations are only of an abstract nature in our minds, and are not at all identical with the images we form visually and through our sense of touch. The non-divisibility of the four-dimensional continuum of events does not at all, however, involve the equivalence of the space co-ordinates with the time co-ordinate. On the contrary, we must remember that the time co-ordinate is defined physically wholly differently from the space co-ordinates. The relations (22) and (22a) which, when equated define the Lorentz transformation show, further, a difference in the rôle of the time co-ordinate from that of the space co-ordinates; for the term Δt^2 has the opposite sign to the space terms, Δx_1^2, Δx_2^2, Δx_3^2.

Before we analyse further the conditions which define the Lorentz transformation, we shall introduce the light-time, $l = ct$, in place of the time, t, in order that the constant c shall not enter explicitly into the formulas to be developed later. Then the Lorentz transformation is defined in such a way that, first, it makes the equation

$$\Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 - \Delta l^2 = 0 \quad . \quad (22b)$$

a co-variant equation, that is, an equation which is satisfied with respect to every inertial system if it is satisfied in the inertial system to which we refer the two given events (emission and reception of the ray of light). Finally, with Minkowski, we introduce in place of the real time co-ordinate $l = ct$, the imaginary time co-ordinate

$$x_4 = il = ict \; (\sqrt{-1} = i)$$

Then the equation defining the propagation of light, which must be co-variant with respect to the Lorentz transformation, becomes

$$\sum_{(4)} \Delta x_\nu^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 + \Delta x_4^2 = 0 \quad . \quad (22c)$$

This condition is always satisfied * if we satisfy the more general condition that

$$s^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 + \Delta x_4^2 \quad . \quad (23)$$

shall be an invariant with respect to the transformation. This condition is satisfied only by linear transformations, that is, transformations of the type

$$x'_\mu = a_\mu + b_{\mu\alpha} x_\alpha \quad . \quad . \quad . \quad (24)$$

in which the summation over the α is to be extended from $\alpha = 1$ to $\alpha = 4$. A glance at equations (23) (24) shows that the Lorentz transformation so defined is identical with the translational and rotational transformations of the Euclidean geometry, if we disregard the number of dimensions and the relations of reality. We can also conclude that the coefficients $b_{\mu\alpha}$ must satisfy the conditions

$$b_{\mu\alpha} b_{\nu\alpha} = \delta_{\mu\nu} = b_{\alpha\mu} b_{\alpha\nu} \quad . \quad . \quad (25)$$

Since the ratios of the x_ν are real, it follows that all the a_μ and the $b_{\mu\alpha}$ are real, except a_4, b_{41}, b_{42}, b_{43}, b_{14}, b_{24}, and b_{34}, which are purely imaginary.

Special Lorentz Transformation. We obtain the simplest transformations of the type of (24) and (25) if only two of the co-ordinates are to be transformed, and if all the a_μ, which merely determine the new origin, vanish. We obtain then for the indices 1 and 2, on account of the three independent conditions which the relations (25) furnish,

$$\left.\begin{aligned} x'_1 &= x_1 \cos\phi - x_2 \sin\phi \\ x'_2 &= x_1 \sin\phi + x_2 \cos\phi \\ x'_3 &= x_3 \\ x'_4 &= x_4 \end{aligned}\right\} \quad . \quad . \quad (26)$$

This is a simple rotation in space of the (space) co-ordinate system about the x_3-axis. We see that the

* That this specialization lies in the nature of the case will be evident later.

rotational transformation in space (without the time transformation) which we studied before is contained in the Lorentz transformation as a special case. For the indices 1 and 4 we obtain, in an analogous manner,

$$\left.\begin{aligned} x'_1 &= x_1 \cos\psi - x_4 \sin\psi \\ x'_4 &= x_1 \sin\psi + x_4 \cos\psi \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned}\right\} \quad . \quad (26a)$$

On account of the relations of reality ψ must be taken as imaginary. To interpret these equations physically, we introduce the real light-time l and the velocity v of K' relatively to K, instead of the imaginary angle ψ. We have, first,

$$\begin{aligned} x'_1 &= x_1 \cos\psi - il \sin\psi \\ l' &= -ix_1 \sin\psi + l \cos\psi \end{aligned}$$

Since for the origin of K', i.e. for $x'_1 = 0$, we must have $x_1 = vl$, it follows from the first of these equations that

$$v = i \tan\psi \quad . \quad . \quad . \quad (27)$$

and also

$$\left.\begin{aligned} \sin\psi &= \frac{-iv}{\sqrt{1 - v^2}} \\ \cos\psi &= \frac{1}{\sqrt{1 - v^2}} \end{aligned}\right\} \quad . \quad . \quad (28)$$

so that we obtain

$$\left.\begin{aligned} x'_1 &= \frac{x_1 - vl}{\sqrt{1 - v^2}} \\ l' &= \frac{l - vx_1}{\sqrt{1 - v^2}} \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned}\right\} \quad . \quad . \quad . \quad (29)$$

These equations form the well-known special Lorentz transformation, which in the general theory represents a rotation, through an imaginary angle, of the four-dimensional system of co-ordinates. If we introduce the ordinary time t, in place of the light-time l, then in (29) we must replace l by ct and v by $\frac{v}{c}$.

We must now fill in a gap. From the principle of the constancy of the velocity of light it follows that the equation

$$\Sigma \Delta x_\nu^2 = 0$$

has a significance which is independent of the choice of the inertial system; but the invariance of the quantity $\Sigma \Delta x_\nu^2$ does not at all follow from this. This quantity might be transformed with a factor. This depends upon the fact that the right-hand side of (29) might be multiplied by a factor λ, which may depend on v. But the principle of relativity does not permit this factor to be different from 1, as we shall now show. Let us assume that we have a rigid circular cylinder moving in the direction of its axis. If its radius measured at rest with a unit measuring-rod is equal to R_0, its radius R in motion might be different from R_0, since the theory of relativity does not make the assumption that the shape of bodies with respect to a space of reference is independent of their motion relatively to this space of reference. But all directions in space must be equivalent to each other. R may therefore depend upon the magnitude q of the velocity, but not upon its direction; R must therefore be an even function of q. If the cylinder is at rest relatively to K' the equation of its lateral surface is

$$x'^2 + y'^2 = R_0^2$$

If we write the last two equations of (29) more generally

$$x'_2 = \lambda x_2$$
$$x'_3 = \lambda x_3$$

then the lateral surface of the cylinder referred to K satisfies the equation

$$x^2 + y^2 = \frac{R_0^2}{\lambda_2}$$

The factor λ therefore measures the lateral contraction of the cylinder, and can thus, from the above, be only an even function of v.

If we introduce a third system of co-ordinates, K'', which moves relatively to K' with velocity v in the direction of the negative x-axis of K, we obtain, by applying (29) twice,

$$x''_1 = \lambda(v)\lambda(-v)x_1$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$l'' = \lambda(v)\lambda(-v)l$$

Now, since $\lambda(v)$ must be equal to $\lambda(-v)$, and since we assume that we use the same measuring-rods in all the systems, it follows that the transformation of K'' to K must be the identical transformation (since the possibility $\lambda = -1$ does not need to be considered). It is essential for these considerations to assume that the behaviour of the measuring-rods does not depend upon the history of their previous motion.

Moving Measuring-Rods and Clocks. At the definite K time, $l = 0$, the position of the points given by the integers $x'_1 = n$, is with respect to K, given by $x_1 = n\sqrt{1 - v^2}$; this follows from the first of equations (29) and expresses the Lorentz contraction. A clock at rest at the origin $x_1 = 0$ of K, whose beats are characterized by $l = n$, will, when observed from K', have beats characterized by

$$l' = \frac{n}{\sqrt{1 - v^2}}$$

this follows from the second of equations (29) and shows that the clock goes slower than if it were at rest relatively to K'. These two consequences, which hold, *mutatis mutandis*, for every system of reference, form the physical content, free from convention, of the Lorentz transformation.

Addition Theorem for Velocities. If we combine two special Lorentz transformations with the relative velocities v_1 and v_2, then the velocity of the single Lorentz transformation which takes the place of the two separate ones is, according to (27), given by

$$v_{12} = i \tan(\psi_1 + \psi_2) = i \frac{\tan\psi_1 + \tan\psi_2}{1 - \tan\psi_1 \tan\psi_2} = \frac{v_1 + v_2}{1 + v_1 v_2} \quad (30)$$

General Statements about the Lorentz Transformation and its Theory of Invariants. The whole theory of invariants of the special theory of relativity depends upon the invariant s^2 (23). Formally, it has the same rôle in the four-dimensional space-time continuum as the invariant $\Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2$ in the Euclidean geometry and in the pre-relativity physics. The latter quantity is not an invariant with respect to all the Lorentz transformations; the quantity s^2 of equation (23) assumes the rôle of this invariant. With respect to an arbitrary inertial system, s^2 may be determined by measurements; with a given unit of measure it is a completely determinate quantity, associated with an arbitrary pair of events.

The invariant s^2 differs, disregarding the number of dimensions, from the corresponding invariant of the Euclidean geometry in the following points. In the Euclidean geometry s^2 is necessarily positive; it vanishes only when the two points concerned come together. On the other hand, from the vanishing of

$$s^2 = \sum \Delta x_\nu^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 - \Delta t^2$$

it cannot be concluded that the two space-time points fall together; the vanishing of this quantity s^2, is the invariant condition that the two space-time points can be connected by a light signal *in vacuo*. If P is a point (event) represented in the four-dimensional space of the x_1, x_2, x_3, l, then all the 'points' which can be connected to P by means of a light signal lie upon the cone $s^2 = 0$ (compare Fig. 1, in which the dimension x_3 is suppressed).

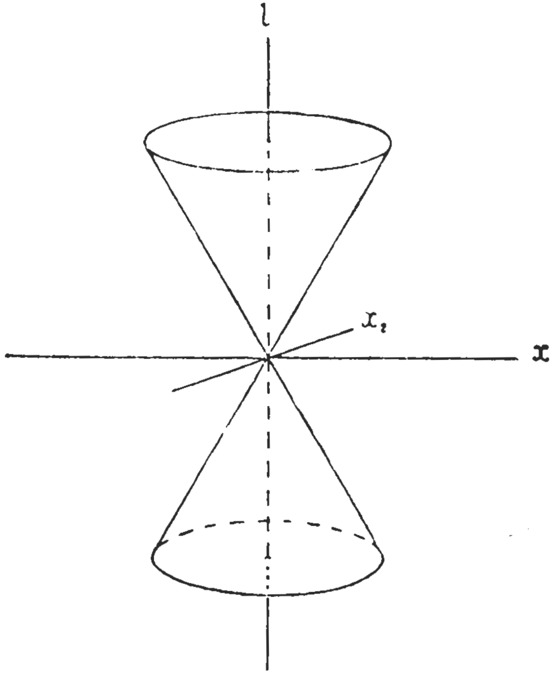

FIG. 1

The 'upper' half of the cone may contain the 'points' to which light signals can be sent from P; then the 'lower' half of the cone will contain the 'points' from which light signals can be sent to P. The points P' enclosed by the conical surface furnish, with P, a negative s^2;

PP', as well as $P'P$ is then, according to Minkowski, of the nature of a time. Such intervals represent elements of possible paths of motion, the velocity being less than that of light.* In this case the l-axis may be drawn in the direction of PP' by suitably choosing the state of motion of the inertial system. If P' lies outside of the 'light-cone' then PP' is of the nature of a space; in this case, by properly choosing the inertial system, Δl can be made to vanish.

By the introduction of the imaginary time variable, $x_4 = il$, Minkowski has made the theory of invariants for the four-dimensional continuum of physical phenomena fully analogous to the theory of invariants for the three-dimensional continuum of Euclidean space. The theory of four-dimensional tensors of special relativity differs from the theory of tensors in three-dimensional space, therefore, only in the number of dimensions and the relations of reality.

A physical entity which is specified by four quantities, A_ν, in an arbitrary inertial system of the x_1, x_2, x_3, x_4, is called a 4-vector, with the components A_ν, if the A_ν correspond in their relations of reality and the properties of transformation to the Δx_ν; it may be space-like or time-like. The sixteen quantities $A_{\mu\nu}$ then form the components of a tensor of the second rank, if they transform according to the scheme

$$A'_{\mu\nu} = b_{\mu\alpha} b_{\nu\beta} A_{\alpha\beta}$$

It follows from this that the $A_{\mu\nu}$ behave, with respect to their properties of transformation and their properties of reality, as the products of the components, U_μ, V_ν, of two 4-vectors, (U) and (V). All the components are real

* That material velocities exceeding that of light are not possible, follows from the appearance of the radical $\sqrt{1 - v^2}$ in the special Lorentz transformation (29).

4-vector) of velocity. Its components satisfy, by (38), the condition

$$\Sigma u_\sigma^2 = -1 \quad . \quad . \quad . \quad (40)$$

We see that this 4-vector, whose components in the ordinary notation are

$$\frac{q_x}{\sqrt{1-q^2}}, \frac{q_y}{\sqrt{1-q^2}}, \frac{q_z}{\sqrt{1-q^2}}, \frac{i}{\sqrt{1-q^2}} \quad . \quad (41)$$

is the only 4-vector which can be formed from the velocity components of the material particle which are defined in three dimensions by

$$q_x = \frac{dx}{dl}, \; q_y = \frac{dy}{dl}, \; q_z = \frac{dz}{dl}$$

We therefore see that

$$\left(m\frac{dx_\mu}{d\tau}\right) \quad . \quad . \quad . \quad . \quad (42)$$

must be that 4-vector which is to be equated to the 4-vector of momentum and energy whose existence we have proved above. By equating the components, we obtain, in three-dimensional notation,

$$\left.\begin{aligned} I_x &= \frac{mq_x}{\sqrt{1-q^2}} \\ &\cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \\ E &= \frac{m}{\sqrt{1-q^2}} \end{aligned}\right\} \quad . \quad . \quad . \quad (43)$$

We recognize, in fact, that these components of momentum agree with those of classical mechanics for velocities which are small compared to that of light. For large velocities the momentum increases more rapidly than linearly with the velocity, so as to become infinite on approaching the velocity of light.

If we apply the last of equations (43) to a material particle at rest ($q = 0$), we see that the energy, E_0, of a body at rest is equal to its mass. Had we chosen the second as our unit of time, we would have obtained

$$E_0 = mc^2 \qquad . \quad . \quad . \quad (44)$$

Mass and energy are therefore essentially alike; they are only different expressions for the same thing. The mass of a body is not a constant; it varies with changes in its energy.* We see from the last of equations (43) that E becomes infinite when q approaches 1, the velocity of light. If we develop E in powers of q^2, we obtain,

$$E = m + \frac{m}{2}q^2 + \frac{3}{8}mq^4 + \ldots \qquad . \quad (45)$$

The second term of this expansion corresponds to the kinetic energy of the material particle in classical mechanics.

Equations of Motion of Material Particles. From (43) we obtain, by differentiating by the time l, and using the principle of momentum, in the notation of three-dimensional vectors,

$$\mathbf{K} = \frac{d}{dl}\left(\frac{mq}{\sqrt{1 - q^2}}\right) \qquad . \quad . \quad . \quad (46)$$

This equation, which was previously employed by H. A. Lorentz for the motion of electrons, has been proved to be true, with great accuracy, by experiments with β-rays.

Energy Tensor of the Electromagnetic Field. Before the development of the theory of relativity it was known that

* The emission of energy in radio-active processes is evidently connected with the fact that the atomic weights are not integers. The equivalence between mass at rest and energy at rest which is expressed in equation (44) has been confirmed in many cases during recent years. In radio-active decomposition the sum of the resulting masses is always less than the mass of the decomposing atom. The difference appears in the form of kinetic energy of the generated particles as well as in the form of released radiational energy.

THE GENERAL THEORY OF RELATIVITY

ALL of the previous considerations have been based upon the assumption that all inertial systems are equivalent for the description of physical phenomena, but that they are preferred, for the formulation of the laws of nature, to spaces of reference in a different state of motion. We can think of no cause for this preference for definite states of motion to all others, according to our previous considerations, either in the perceptible bodies or in the concept of motion; on the contrary, it must be regarded as an independent property of the space-time continuum. The principle of inertia, in particular, seems to compel us to ascribe physically objective properties to the space-time continuum. Just as it was consistent from the Newtonian standpoint to make both the statements, *tempus est absolutum, spatium est absolutum*, so from the standpoint of the special theory of relativity we must say, *continuum spatii et temporis est absolutum*. In this latter statement *absolutum* means not only 'physically real', but also 'independent in its physical properties, having a physical effect, but not itself influenced by physical conditions'.

As long as the principle of inertia is regarded as the keystone of physics, this standpoint is certainly the only one which is justified. But there are two serious criticisms of the ordinary conception. In the first place, it is contrary to the mode of thinking in science to conceive of a thing (the space-time continuum) which acts itself, but which cannot be acted upon. This is the reason why E. Mach was led to make the attempt to eliminate space as an active cause in the system of mechanics. According to him, a material particle does not move in unaccelerated motion relatively to space, but relatively to the centre of all the other masses in the universe ; in this way the series of causes of mechanical phenomena was closed, in contrast to the mechanics of Newton and Galileo. In order to develop this idea within the limits of the modern theory of action through a medium, the properties of the space-time continuum which determine inertia must be regarded as field properties of space, analogous to the electromagnetic field. The concepts of classical mechanics afford no way of expressing this. For this reason Mach's attempt at a solution failed for the time being. We shall come back to this point of view later. In the second place, classical mechanics exhibits a deficiency which directly calls for an extension of the principle of relativity to spaces of reference which are not in uniform motion relatively to each other. The ratio of the masses of two bodies is defined in mechanics in two ways which differ from each other fundamentally ; in the first place, as the reciprocal ratio of the accelerations which the same motive force imparts to them (inert mass), and in the second place, as the ratio of the forces which act upon them in the same gravitational field (gravitational mass). The equality of these two masses, so differently defined, is a fact which is confirmed by experiments of very high accuracy (experiments of Eötvös), and classical mechanics offers no explanation for this equality. It is, however, clear that science is fully justified in assigning such a numerical equality only after this numerical equality is reduced to an equality of the real nature of the two concepts.

That this object may actually be attained by an extension of the principle of relativity, follows from the following consideration. A little reflection will show that the law of the equality of the inert and the gravitational mass is equivalent to the assertion that the acceleration imparted to a body by a gravitational field is independent of the nature of the body. For Newton's equation of motion in a gravitational field, written out in full, is

(Inert mass) . (Acceleration) = (Intensity of the gravitational field) . (Gravitational mass)

It is only when there is numerical equality between the inert and gravitational mass that the acceleration is independent of the nature of the body. Let now K be an inertial system. Masses which are sufficiently far from each other and from other bodies are then, with respect to K, free from acceleration. We shall also refer these masses to a system of co-ordinates K', uniformly accelerated with respect to K. Relatively to K' all the masses have equal and parallel accelerations; with respect to K' they behave just as if a gravitational field were present and K' were unaccelerated. Overlooking for the present the question as to the 'cause' of such a gravitational field, which will occupy us later, there is nothing to prevent our conceiving this gravitational field as real, that is, the conception that K' is 'at rest' and a gravitational field is present we may consider as equivalent to the conception that only K is an 'allowable' system of co-ordinates and no gravitational field is present. The assumption of the complete physical equivalence of the systems of co-ordinates, K and K', we call the 'principle of equivalence'; this principle is evidently intimately connected with the law of the equality between the inert and the gravitational mass, and signifies an extension of the principle of relativity to co-ordinate systems which are in non-uniform motion relatively to each other. In fact, through this conception we arrive at the unity of the nature of inertia and gravitation. For, according to our way of looking at it, the same masses may appear to be either under the action of inertia alone (with respect to K) or under the combined action of inertia and gravitation (with respect to K'). The possibility of explaining the numerical equality of inertia and gravitation by the unity of their nature gives to the general theory of relativity, according to my conviction, such a superiority over the conceptions of classical mechanics, that all the difficulties encountered in development must be considered as small in comparison with this progress.

What justifies us in dispensing with the preference for inertial systems over all other co-ordinate systems, a preference that seems so securely established by experience? The weakness of the principle of inertia lies in this, that it involves an argument in a circle: a mass moves without acceleration if it is sufficiently far from other bodies; we know that it is sufficiently far from other bodies only by the fact that it moves without acceleration. Are there, at all, any inertial systems for very extended portions of the space-time continuum, or, indeed, for the whole universe? We may look upon the principle of inertia as established, to a high degree of approximation, for the space of our planetary system, provided that we neglect the perturbations due to the sun and planets. Stated more exactly, there are finite regions, where, with respect to a suitably chosen space of reference, material particles move freely without acceleration, and in which the laws of the special theory of relativity, which have been developed above, hold with remarkable accuracy. Such regions we shall call 'Galilean regions'. We shall proceed from the consideration of such regions as a special case of known properties.

The principle of equivalence demands that in dealing with Galilean regions we may equally well make use of non-inertial systems, that is, such co-ordinate systems as, relatively to inertial systems, are not free from acceleration and rotation. If, further, we are going to do away completely with the vexing question as to the objective reason for the preference of certain systems of co-ordinates, then we must allow the use of arbitrarily moving systems of co-ordinates. As soon as we make this attempt seriously we come into conflict with that physical interpretation of space and time to which we were led by the special theory of relativity. For let K' be a system of co-ordinates whose z'-axis coincides with the z-axis of K, and which rotates about the latter axis with constant angular velocity. Are the configurations of rigid bodies, at rest relatively to K', in accordance with the laws of Euclidean geometry? Since K' is not an inertial system, we do not know directly the laws of configuration of rigid bodies with respect to K', nor the laws of nature, in general. But we do know these laws with respect to the inertial system K, and we can therefore infer their form with respect to K'. Imagine a circle drawn about the origin in the $x'y'$ plane of K', and a diameter of this circle. Imagine, further, that we have given a large number of rigid rods, all equal to each other. We suppose these laid in series along the periphery and the diameter of the circle, at rest relatively to K'. If U is the number of these rods along the periphery, D the number along the diameter, then, if K' does not rotate relatively to K, we shall have

$$\frac{U}{D} = \pi$$

But if K' rotates we get a different result. Suppose that at a definite time t, of K we determine the ends of all the rods. With respect to K all the rods upon the periphery experience the Lorentz contraction, but the rods upon the diameter do not experience this contraction (along their lengths !).* It therefore follows that

$$\frac{U}{D} > \pi$$

It therefore follows that the laws of configuration of rigid bodies with respect to K' do not agree with the laws of configuration of rigid bodies that are in accordance with Euclidean geometry. If, further, we place two similar clocks (rotating with K'), one upon the periphery, and the other at the centre of the circle, then, judged from K, the clock on the periphery will go slower than the clock at the centre. The same thing must take place, judged from K', if we do not define time with respect to K' in a wholly unnatural way (that is, in such a way that the laws with respect to K' depend explicitly upon the time). Space and time, therefore, cannot be defined with respect to K' as they were in the special theory of relativity with respect to inertial systems. But, according to the principle of equivalence, K' may also be considered as a system at rest, with respect to which there is a gravitational field (field of centrifugal force, and force of Coriolis). We therefore arrive at the result : the gravitational field influences and even determines the metrical laws of the space-time continuum. If the laws of configuration of ideal rigid bodies are to be expressed geometrically, then in the presence of a gravitational field the geometry is not Euclidean.

The case that we have been considering is analogous to that which is presented in the two-dimensional treatment of surfaces. It is impossible in the latter case also, to introduce co-ordinates on a surface (e.g. the surface of

* These considerations assume that the behaviour of rods and clocks depends only upon velocities, and not upon accelerations, or, at least, that the influence of acceleration does not counteract that of velocity.

an ellipsoid) which have a simple metrical significance, while on a plane the Cartesian co-ordinates, x_1, x_2, signify directly lengths measured by a unit measuring-rod. Gauss overcame this difficulty, in his theory of surfaces, by introducing curvilinear co-ordinates which, apart from satisfying conditions of continuity, were wholly arbitrary, and only afterwards were these co-ordinates related to the metrical properties of the surface. In an analogous way we shall introduce in the general theory of relativity arbitrary co-ordinates, x_1, x_2, x_3, x_4, which shall number uniquely the space-time points, so that neighbouring events are associated with neighbouring values of the co-ordinates; otherwise, the choice of co-ordinates is arbitrary. We shall be true to the principle of relativity in its broadest sense if we give such a form to the laws that they are valid in every such four-dimensional system of co-ordinates, that is, if the equations expressing the laws are co-variant with respect to arbitrary transformations.

The most important point of contact between Gauss's theory of surfaces and the general theory of relativity lies in the metrical properties upon which the concepts of both theories, in the main, are based. In the case of the theory of surfaces, Gauss's argument is as follows. Plane geometry may be based upon the concept of the distance ds, between two infinitely near points. The concept of this distance is physically significant because the distance can be measured directly by means of a rigid measuring-rod. By a suitable choice of Cartesian co-ordinates this distance may be expressed by the formula $ds^2 = dx_1{}^2 + dx_2{}^2$. We may base upon this quantity the concepts of the straight line as the geodesic $\left(\delta\int ds = 0\right)$, the interval, the circle, and the angle, upon which the Euclidean plane geometry is built. A geometry may be developed upon another continuously curved surface, if we observe that an infinitesimally small portion of the surface may be regarded as plane, to within relatively infinitesimal quantities. There are Cartesian co-ordinates, X_1, X_2, upon such a small portion of the surface, and the distance between two points, measured by a measuring-rod, is given by

$$ds^2 = dX_1^2 + dX_2^2$$

If we introduce arbitrary curvilinear co-ordinates, x_1, x_2, on the surface, then dX_1, dX_2, may be expressed linearly in terms of dx_1, dx_2. Then everywhere upon the surface we have

$$ds^2 = g_{11}\, dx_1^2 + 2g_{12}\, dx_1\, dx_2 + g_{22}\, dx_2^2$$

where g_{11}, g_{12}, g_{22} are determined by the nature of the surface and the choice of co-ordinates; if these quantities are known, then it is also known how networks of rigid rods may be laid upon the surface. In other words, the geometry of surfaces may be based upon this expression for ds^2 exactly as plane geometry is based upon the corresponding expression.

There are analogous relations in the four-dimensional space-time continuum of physics. In the immediate neighbourhood of an observer, falling freely in a gravitational field, there exists no gravitational field. We can therefore always regard an infinitesimally small region of the space-time continuum as Galilean. For such an infinitely small region there will be an inertial system (with the space co-ordinates, X_1, X_2, X_3, and the time co-ordinate X_4) relatively to which we are to regard the laws of the special theory of relativity as valid. The quantity which is directly measurable by our unit measuring-rods and clocks,

$$dX_1^2 + dX_2^2 + dX_3^2 - dX_4^2$$

or its negative,

$$ds^2 = -dX_1^2 - dX_2^2 - dX_3^2 + dX_4^2 \quad . \quad (54)$$

is therefore a uniquely determinate invariant for two neighbouring events (points in the four-dimensional continuum), provided that we use measuring-rods that are equal to each other when brought together and superimposed, and clocks whose rates are the same when they are brought together. In this the physical assumption is essential that the relative lengths of two measuring-rods and the relative rates of two clocks are independent, in principle, of their previous history. But this assumption is certainly warranted by experience; if it did not hold there could be no sharp spectral lines, since the single atoms of the same element certainly do not have the same listing, and since—on the assumption of relative variability of the single atoms depending on previous history—it would be absurd to suppose that the masses or proper frequencies of these atoms ever had been equal to one another.

Space-time regions of finite extent are, in general, not Galilean, so that a gravitational field cannot be done away with by any choice of co-ordinates in a finite region. There is, therefore, no choice of co-ordinates for which the metrical relations of the special theory of relativity hold in a finite region. But the invariant ds always exists for two neighbouring points (events) of the continuum. This invariant ds may be expressed in arbitrary co-ordinates. If one observes that the local dX_ν may be expressed linearly in terms of the co-ordinate differentials dx_ν, ds^2 may be expressed in the form

$$ds^2 = g_{\mu\nu}\, dx_\mu\, dx_\nu \quad . \quad . \quad . \quad (55)$$

The functions $g_{\mu\nu}$ describe, with respect to the arbitrarily chosen system of co-ordinates, the metrical relations of the space-time continuum and also the gravitational field. As in the special theory of relativity, we have to discriminate between time-like and space-like line elements in the four-dimensional continuum ; owing to the change of sign introduced, time-like line elements have a real, space-like line elements an imaginary *ds*. The time-like *ds* can be measured directly by a suitably chosen clock.

According to what has been said, it is evident that the formulation of the general theory of relativity requires a generalization of the theory of invariants and the theory of tensors ; the question is raised as to the form of the equations which are co-variant with respect to arbitrary point transformations. The generalized calculus of tensors was developed by mathematicians long before the theory of relativity. Riemann first extended Gauss's train of thought to continua of any number of dimensions ; with prophetic vision he saw the physical meaning of this generalization of Euclid's geometry. Then followed the development of the theory in the form of the calculus of tensors, particularly by Ricci and Levi-Civita. This is the place for a brief presentation of the most important mathematical concepts and operations of this calculus of tensors.

We designate four quantities, which are defined as functions of the x_ν with respect to every system of co-ordinates, as components, A^ν, of a contra-variant vector, if they transform in a change of co-ordinates as the co-ordinate differentials dx_ν. We therefore have

$$A^{\mu'} = \frac{\partial x'_{\mu}}{\partial x_\nu} A \quad . \quad . \quad . \quad (56)$$

Besides these contra-variant vectors, there are also co-variant vectors. If B_ν are the components of a co-variant vector, these vectors are transformed according to the rule

$$B'_\mu = \frac{dx_\nu}{\partial x'_\mu} B_\nu \quad . \quad . \quad . \quad (57)$$

The definition of a co-variant vector is chosen in such

Δx_2, Δx_3, corresponding to the ends of a unit measuring-rod, oriented in any way, shall always satisfy the relation $\Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 = 1$. In this sense space is not Euclidean, but 'curved'. It follows from the second of the relations above that the interval between two beats of the unit clock ($dT = 1$) corresponds to the 'time'

$$1 + \frac{\kappa}{8\pi}\int\frac{\sigma dV_0}{r}$$

in the unit used in our system of co-ordinates. The rate of a clock is accordingly slower the greater is the mass of the ponderable matter in its neighbourhood. We therefore conclude that spectral lines which are produced on the sun's surface will be displaced towards the red, compared to the corresponding lines produced on the earth, by about $2 \cdot 10^{-6}$ of their wave-lengths. At first, this important consequence of the theory appeared to conflict with experiment; but results obtained during the past years seem to make the existence of this effect more and more probable, and it can hardly be doubted that this consequence of the theory will be confirmed within the next years.

Another important consequence of the theory, which can be tested experimentally, has to do with the path of rays of light. In the general theory of relativity also the velocity of light is everywhere the same, relatively to a local inertial system. This velocity is unity in our natural measure of time. The law of the propagation of light in general co-ordinates is therefore, according to the general theory of relativity, characterized, by the equation

$$ds^2 = 0$$

To within the approximation which we are using, and in the system of co-ordinates which we have selected, the velocity of light is characterized, according to (106), by the equation

$$\left(1 + \frac{\kappa}{4\pi}\int\frac{\sigma dV_0}{r}\right)(dx_1^2 + dx_2^2 + dx_3^2) = \left(1 - \frac{\kappa}{4\pi}\int\frac{\sigma dV_0}{r}\right)dl^2$$

The velocity of light L, is therefore expressed in our co-ordinates by

$$\frac{\sqrt{dx_1^2 + dx_2^2 + dx_3^2}}{dl} = 1 - \frac{\kappa}{4\pi}\int\frac{\sigma dV_0}{r} \quad . \quad (107)$$

We can therefore draw the conclusion from this, that a ray of light passing near a large mass is deflected. If we imagine the sun, of mass M, concentrated at the origin of our system of co-ordinates, then a ray of light, travelling parallel to the x_3-axis, in the $x_1 - x_3$ plane, at a distance Δ from the origin, will be deflected, in all, by an amount

$$\alpha = \int_{-\infty}^{+\infty} \frac{1}{L}\frac{\partial L}{\partial x_1} dx_3$$

towards the sun. On performing the integration we get

$$\alpha = \frac{\kappa M}{2\pi\Delta} \quad . \quad . \quad . \quad (108)$$

The existence of this deflection, which amounts to 1.7′ for Δ equal to the radius of the sun, was confirmed, with remarkable accuracy, by the English Solar Eclipse Expedition in 1919, and most careful preparations have been made to get more exact observational data at the solar eclipse in 1922. It should be noted that this result, also, of the theory is not influenced by our arbitrary choice of a system of co-ordinates.

This is the place to speak of the third consequence of the theory which can be tested by observation, namely, that which concerns the motion of the perihelion of the planet Mercury. The secular changes in the planetary orbits are known with such accuracy that the approximation we have been using is no longer sufficient for a comparison of theory and observation. It is necessary to go back to the general field equations (96). To solve this problem I made use of the method of successive approximations. Since then, however, the problem of the central symmetrical statical gravitational field has been completely solved by Schwarzschild and others; the derivation given by H. Weyl in his book, *Raum-Zeit-Materie*, is particularly elegant. The calculation can be simplified somewhat if we do not go back directly to the equation (96), but base it upon a principle of variation that is equivalent to this equation. I shall indicate the procedure only in so far as is necessary for understanding the method.

In the case of a statical field, ds^2 must have the form

$$\left.\begin{aligned} ds^2 &= -d\sigma^2 + f^2\,dx_4^2 \\ d\sigma^2 &= \sum_{1-3}\gamma_{\alpha\beta}\,dx_\alpha\,dx_\beta \end{aligned}\right\} \quad . \quad . \quad (109)$$

where the summation on the right-hand side of the last equation is to be extended over the space variables only. The central symmetry of the field requires the $\gamma_{\mu\nu}$ to be of the form,

$$\gamma_{\alpha\beta} = \mu\delta_{\alpha\beta} + \lambda x_\alpha x_\beta \quad . \quad . \quad (110)$$

f^2, μ and λ are functions of $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$ only. One of these three functions can be chosen arbitrarily, because our system of co-ordinates is, *a priori*, completely arbitrary; for by a substitution

$$\begin{aligned} x'_4 &= x_4 \\ x'_\alpha &= F(r)x_\alpha \end{aligned}$$

we can always insure that one of these three functions shall be an assigned function of r'. In place of (110) we can therefore put, without limiting the generality

$$\gamma_{\alpha\beta} = \delta_{\alpha\beta} + \lambda x_\alpha x_\beta \quad . \quad . \quad (110a)$$

In this way the $g_{\mu\nu}$ are expressed in terms of the two quantities λ and f. These are to be determined as functions of r, by introducing them into equation (96), after first calculating the $\Gamma^{\sigma}_{\mu\nu}$ from (109) and (110a). We have

$$\left.\begin{aligned} \Gamma^{\sigma}_{\alpha\beta} &= \frac{1}{2}\frac{x_\sigma}{r} \cdot \frac{x_\alpha x_\beta + 2\lambda r \delta_{\alpha\beta}}{1 + \lambda r^2} \text{ (for } \alpha, \beta, \sigma = 1, 2, 3) \\ \Gamma^{4}_{44} &= \Gamma^{\alpha}_{4\beta} = \Gamma^{4}_{\alpha\beta} = 0 \text{ (for } \alpha, \beta = 1, 2, 3) \\ \Gamma^{4}_{4\alpha} &= \frac{1}{2} f^{-2} \frac{\partial f^2}{\partial x_\alpha}, \quad \Gamma^{\alpha}_{44} = -\frac{1}{2} g^{\alpha\beta} \frac{\partial f^2}{\partial x_\beta} \end{aligned}\right\} \qquad (110b)$$

With the help of these results, the field equations furnish Schwarzschild's solution :

$$ds^2 = \left(1 - \frac{A}{r}\right) dl^2 - \left[\frac{dr^2}{1 - \frac{A}{r}} + r^2(\sin^2\theta \, d\phi^2 + d\theta^2)\right] \qquad (109a)$$

in which we have put

$$\left.\begin{aligned} x_4 &= l \\ x_1 &= r \sin\theta \sin\phi \\ x_2 &= r \sin\theta \cos\phi \\ x_3 &= r \cos\theta \\ A &= \frac{\kappa M}{4\pi} \end{aligned}\right\} \quad . \quad . \quad (109b)$$

M denotes the sun's mass, centrally symmetrically placed about the origin of co-ordinates ; the solution (109) is valid only outside of this mass, where all the $T_{\mu\nu}$ vanish. If the motion of the planet takes place in the $x_1 - x_2$ plane then we must replace (109a) by

$$ds^2 = \left(1 - \frac{A}{r}\right) dl^2 - \frac{dr^2}{1 - \frac{A}{r}} - r^2 \, d\phi^2 \quad . \quad (109c)$$

The calculation of the planetary motion depends upon equation (90). From the first of equations (110b) and (90) we get, for the indices, 1, 2, 3,

$$\frac{d}{ds}\left(x_\alpha \frac{dx_\beta}{ds} - x_\beta \frac{dx_\alpha}{ds}\right) = 0$$

or, if we integrate, and express the result in polar co-ordinates,

$$r^2 \frac{d\phi}{ds} = \text{constant} \qquad . \qquad . \qquad (111)$$

From (90), for $\mu = 4$, we get

$$0 = \frac{d^2 l}{ds^2} + \frac{1}{f^2}\frac{\partial f^2}{\partial x_\alpha}\frac{dx^\alpha}{ds}\frac{dl}{ds} = \frac{d^2 l}{ds^2} + \frac{1}{f^2}\frac{df^2}{ds}\frac{dl}{ds}$$

From this, after multiplication by f^2 and integration, we have

$$f^2 \frac{dl}{ds} = \text{constant} \qquad . \qquad . \qquad (112)$$

In (109c), (111) and (112) we have three equations between the four variables s, r, l and ϕ, from which the motion of the planet may be calculated in the same way as in classical mechanics. The most important result we get from this is a secular rotation of the elliptic orbit of the planet in the same sense as the revolution of the planet, amounting in radians per revolution to

$$\frac{24\pi^3 a^2}{(1 - e^2)c^2 T^2} \qquad . \qquad . \qquad . \qquad (113)$$

where a = the semi-major axis of the planetary orbit in centimetres.

e = the numerical eccentricity.

$c = 3 \cdot 10^{+10}$, the velocity of the light *in vacuo*.

T = the period of revolution in seconds.

This expression furnishes the explanation of the motion of the perihelion of the planet Mercury, which has been known for a hundred years (since Leverrier), and for which theoretical astronomy has hitherto been unable satisfactorily to account.

There is no difficulty in expressing Maxwell's theory of the electromagnetic field in terms of the general theory of relativity; this is done by application of the tensor formation (81), (82) and (77). Let ϕ_μ be a tensor of the first rank, to be interpreted as an electromagnetic 4-potential; then an electromagnetic field tensor may be defined by the relations,

$$\phi_{\mu\nu} = \frac{\partial \phi_\mu}{\partial x_\mu} - \frac{\partial \phi_\nu}{\partial x_\mu} \quad . \quad . \quad (114)$$

The second of Maxwell's systems of equations is then defined by the tensor equation, resulting from this,

$$\frac{\partial \phi_{\mu\nu}}{\partial x_\rho} + \frac{\partial \phi_{\nu\rho}}{\partial x_\mu} + \frac{\partial \phi_{\rho\mu}}{\partial x_\nu} = 0 \quad . \quad . \quad (114a)$$

and the first of Maxwell's systems of equations is defined by the tensor-density relation

$$\frac{\partial \mathfrak{F}^{\mu\nu}}{\partial x_\nu} = \mathfrak{J}^\mu \quad . \quad . \quad . \quad (115)$$

in which

$$\mathfrak{F}^{\mu\nu} = \sqrt{-g}\, g^{\mu\sigma} g^{\nu\tau} \phi_{\sigma\tau}$$

$$\mathfrak{J}^\mu = \sqrt{-g}\, \rho \frac{dx_\nu}{ds}$$

If we introduce the energy tensor of the electromagnetic field into the right-hand side of (96), we obtain (115), for the special case $\mathfrak{J}^\mu = 0$, as a consequence of (96) by taking the divergence. This inclusion of the theory of electricity in the scheme of the general theory of relativity has been considered arbitrary and unsatisfactory by many theoreticians. Nor can we in this way understand the equilibrium of the electricity which constitutes the elementary electrically charged particles. A theory in which the gravitational field and the electromagnetic field do not enter as logically distinct structures would be much preferable. H. Weyl, and recently Th. Kaluza, have put forward ingenious ideas along this direction; but concerning them, I am convinced that they do not bring us nearer to the true solution of the fundamental problem. I shall not go into this further, but shall give a brief discussion of the so-called cosmological problem, for without this, the considerations regarding the general theory of relativity would, in a certain sense, remain unsatisfactory.

Our previous considerations, based upon the field equations (96), had for a foundation the conception that space on the whole is Galilean-Euclidean, and that this character is disturbed only by masses embedded in it. This conception was certainly justified as long as we were dealing with spaces of the order of magnitude of those that astronomy has mostly to do with. But whether portions of the universe, however large they may be, are quasi-Euclidean, is a wholly different question. We can make this clear by using an example from the theory of surfaces which we have employed many times. If a certain portion of a surface appears to be practically plane, it does not at all follow that the whole surface has the form of a plane; the surface might just as well be a sphere of sufficiently large radius. The question as to whether the universe as a whole is non-Euclidean was much discussed from the geometrical point of view before the development of the theory of relativity. But with the theory of relativity, this problem has entered upon a new stage, for according to this theory the geometrical properties of bodies are not independent, but depend upon the distribution of masses.

If the universe were quasi-Euclidean, then Mach was wholly wrong in his thought that inertia, as well as gravitation, depends upon a kind of mutual action between bodies. For in this case, for a suitably selected system of co-ordinates, the $g_{\mu\nu}$ would be constant at infinity, as they are in the special theory of relativity, while within finite regions the $g_{\mu\nu}$ would differ from these constant values by small amounts only, for a suitable choice of co-ordinates, as a result of the influence of the masses in finite regions. The physical properties of space would not then be wholly independent, that is uninfluenced by matter, but in the main they would be, and only in small measure, conditioned by matter. Such a dualistic conception is even in itself not satisfactory; there are, however, some important physical arguments against it, which we shall consider.

The hypothesis that the universe is infinite and Euclidean at infinity, is, from the relativistic point of view, a complicated hypothesis. In the language of the general theory of relativity it demands that the Riemann tensor of the fourth rank, R_{iklm}, shall vanish at infinity, which furnishes twenty independent conditions, while only ten curvature components, $R_{\mu\nu}$, enter into the laws of the gravitational field. It is certainly unsatisfactory to postulate such a far-reaching limitation without any physical basis for it.

But in the second place, the theory of relativity makes it appear probable that Mach was on the right road in his thought that inertia depends upon a mutual action of matter. For we shall show in the following that, according to our equations, inert masses do act upon each other in the sense of the relativity of inertia, even if only very feebly. What is to be expected along the line of Mach's thought?

1 The inertia of a body must increase when ponderable masses are piled up in its neighbourhood.

2. A body must experience an accelerating force when

neighbouring masses are accelerated, and, in fact, the force must be in the same direction as that acceleration.

3. A rotating hollow body must generate inside of itself a 'Coriolis field', which deflects moving bodies in the sense of the rotation, and a radial centrifugal field as well.

We shall now show that these three effects, which are to be expected in accordance with Mach's ideas, are actually present according to our theory, although their magnitude is so small that confirmation of them by laboratory experiments is not to be thought of. For this purpose we shall go back to the equations of motion of a material particle (90), and carry the approximations somewhat further than was done in equation (90a).

First, we consider γ_{44} as small of the first order. The square of the velocity of masses moving under the influence of the gravitational force is of the same order, according to the energy equation. It is therefore logical to regard the velocities of the material particles we are considering, as well as the velocities of the masses which generate the field, as small, of the order $\frac{1}{2}$. We shall now carry out the approximation in the equations that arise from the field equations (101) and the equations of motion (90) so far as to consider terms, in the second member of (90), that are linear in those velocities. Further, we shall not put ds and dl equal to each other, but, corresponding to the higher approximation, we shall put

$$ds = \sqrt{g_{44}}dl = \left(1 - \frac{\gamma_{44}}{2}\right)dl$$

From (90) we obtain, at first,

$$\frac{d}{dl}\left[\left(1 + \frac{\gamma_{44}}{2}\right)\frac{dx_\mu}{dl}\right] = -\Gamma^\mu_{\alpha\beta}\frac{dx_\alpha}{dl}\frac{dx_\beta}{dl}\left(1 + \frac{\gamma_{44}}{2}\right) \quad (116)$$

From (101) we get, to the approximation sought for,

$$\left.\begin{aligned} -\gamma_{11} = -\gamma_{22} = -\gamma_{33} = \gamma_{44} &= \frac{\kappa}{4\pi}\int\frac{\sigma dV_0}{r} \\ \gamma_{4\alpha} &= -\frac{i\kappa}{2\pi}\int\frac{\sigma\frac{dx_\alpha}{ds}dV_0}{r} \\ \gamma_{\alpha\beta} &= 0 \end{aligned}\right\} \quad (117)$$

in which. in (117), α and β denote the space indices only.

On the right-hand side of (116) we can replace $1 + \frac{\gamma_{44}}{2}$ by 1 and $-\Gamma_\mu^{\alpha\beta}$ by $\left[{\alpha\beta \atop \mu}\right]$. It is easy to see, in addition, that to this degree of approximation we must put

$$\begin{aligned} \left[{44 \atop \mu}\right] &= -\tfrac{1}{2}\frac{\partial\gamma_{44}}{\partial x_\mu} + \frac{\partial\gamma_{4\mu}}{\partial x_4} \\ \left[{\alpha 4 \atop \mu}\right] &= \tfrac{1}{2}\left(\frac{\partial\gamma_{4\mu}}{\partial x_\alpha} - \frac{\partial\gamma_{4\alpha}}{\partial x_\mu}\right) \\ \left[{\alpha\beta \atop \mu}\right] &= 0 \end{aligned}$$

in which α, β and μ denote space indices. We therefore obtain from (116), in the usual vector notation,

$$\left.\begin{aligned} \frac{d}{dl}[(1+\bar\sigma)\mathbf{v}] &= \operatorname{grad}\bar\sigma + \frac{\partial\mathfrak{A}}{\partial l} + [\operatorname{rot}\mathfrak{A}\times\mathbf{v}] \\ \bar\sigma &= \frac{\kappa}{8\pi}\int\frac{\sigma dV_0}{r} \\ \mathfrak{A} &= \frac{\kappa}{2\pi}\int\frac{\sigma\frac{dx_\alpha}{dl}dV_0}{r} \end{aligned}\right\} \quad (118)$$

The equations of motion, (118), show now, in fact, that

1. The inert mass is proportional to $1 + \bar\sigma$, and therefore increases when ponderable masses approach the test body.

2. There is an inductive action of accelerated masses, of the same sign, upon the test body. This is the term $\frac{\partial \mathfrak{A}}{\partial l}$.
3. A material particle, moving perpendicularly to the axis of rotation inside a rotating hollow body, is deflected in the sense of the rotation (Coriolis field). The centrifugal action, mentioned above, inside a rotating hollow body, also follows from the theory, as has been shown by Thirring.*

Although all of these effects are inaccessible to experiment, because κ is so small, nevertheless they certainly exist according to the general theory of relativity. We must see in them a strong support for Mach's ideas as to the relativity of all inertial actions. If we think these ideas consistently through to the end we must expect the *whole* inertia, that is, the *whole* $g_{\mu\nu}$-field, to be determined by the matter of the universe, and not mainly by the boundary conditions at infinity.

For a satisfactory conception of the $g_{\mu\nu}$-field of cosmical dimensions, the fact seems to be of significance that the relative velocity of the stars is small compared to the velocity of light. It follows from this that, with a suitable choice of co-ordinates, g_{44} is nearly constant in the universe, at least, in that part of the universe in which there is matter. The assumption appears natural, moreover, that there are stars in all parts of the universe, so that we may well assume that the inconstancy of g_{44} depends only upon the circumstance that matter is not distributed continuously, but is concentrated in single,

* That the centrifugal action must be inseparably connected with the existence of the Coriolis field may be recognized, even without calculation, in the special case of a co-ordinate system rotating uniformly relatively to an inertial system; our general co-variant equations naturally must apply to such a case.

celestial bodies and systems of bodies. If we are willing to ignore these more local non-uniformities of the density of matter and of the $g_{\mu\nu}$-field, in order to learn something of the geometrical properties of the universe as a whole, it appears natural to substitute for the actual distribution of masses a continuous distribution, and furthermore to assign to this distribution a uniform density σ. In this imagined universe all points with space directions will be geometrically equivalent; with respect to its space extension it will have a constant curvature, and will be cylindrical with respect to its x_4-co-ordinate. The possibility seems to be particularly satisfying that the universe is spatially bounded and thus, in accordance with our assumption of the constancy of σ, is of constant curvature, being either spherical or elliptical; for then the boundary conditions at infinity which are so inconvenient from the standpoint of the general theory of relativity, may be replaced by the much more natural conditions for a closed space.

According to what has been said, we are to put

$$ds^2 = dx_4{}^2 - \gamma_{\mu\nu}dx_\mu dx_\nu \quad . \quad . \quad (119)$$

in which the indices μ and ν run from 1 to 3 only. The $\gamma_{\mu\nu}$ will be such functions of x_1, x_2, x_3 as correspond to a three-dimensional continuum of constant positive curvature. We must now investigate whether such an assumption can satisfy the field equations of gravitation.

In order to be able to investigate this, we must first find what differential conditions the three-dimensional manifold of constant curvature satisfies. A spherical manifold of three dimensions, embedded in a Euclidean continuum of four dimensions,* is given by the equations

$$x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2 = a^2$$
$$dx_1{}^2 + dx_2{}^2 + dx_3{}^2 + dx_4{}^2 = ds^2$$

* The aid of a fourth space dimension has naturally no significance except that of a mathematical artifice.

By eliminating x_4, we get

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1\,dx_1 + x_2\,dx_2 + x_3\,dx_3)^2}{a^2 - x_1^2 - x_2^2 - x_3^2}$$

Neglecting terms of the third and higher degrees in the x_ν, we can put, in the neighbourhood of the origin of co-ordinates,

$$ds^2 = \left(\delta_{\mu\nu} + \frac{x_\mu x_\nu}{a^2}\right) dx_\mu\,dx_\nu$$

Inside the brackets are the $g_{\mu\nu}$ of the manifold in the neighbourhood of the origin. Since the first derivatives of the $g_{\mu\nu}$, and therefore also the $\Gamma^\sigma_{\mu\nu}$, vanish at the origin, the calculation of the $R_{\mu\nu}$ for this manifold, by (88), is very simple at the origin. We have

$$R_{\mu\nu} = -\frac{2}{a^2}\delta_{\mu\nu} = -\frac{2}{a^2}g_{\mu\nu}$$

Since the relation $R_{\mu\nu} = \frac{2}{a^2}g_{\mu\nu}$ is generally co-variant, and since all points of the manifold are geometrically equivalent, this relation holds for every system of co-ordinates, and everywhere in the manifold. In order to avoid confusion with the four-dimensional continuum, we shall, in the following, designate quantities that refer to the three-dimensional continuum by Greek letters, and put

$$\mathrm{P}_{\mu\nu} = -\frac{2}{a^2}\gamma_{\mu\nu} \quad . \quad . \quad (120)$$

We now proceed to apply the field equations (96) to our special case. From (119) we get for the four-dimensional manifold,

$$\left.\begin{array}{l} R_{\mu\nu} = \mathrm{P}_{\mu\nu} \text{ for the indices 1 to 3} \\ R_{14} = R_{24} = R_{34} = R_{44} = 0 \end{array}\right\} \quad . \quad (121)$$

For the right-hand side of (96) we have to consider the energy tensor for matter distributed like a cloud of dust. According to what has gone before we must therefore put

$$T^{\mu\nu} = \sigma \frac{dx_\mu}{ds} \frac{dx_\nu}{ds}$$

specialized for the case of rest. But in addition, we shall add a pressure term that may be physically established as follows. Matter consists of electrically charged particles. On the basis of Maxwell's theory these cannot be conceived of as electromagnetic fields free from singularities. In order to be consistent with the facts, it is necessary to introduce energy terms, not contained in Maxwell's theory, so that the single electric particles may hold together in spite of the mutual repulsions between their elements, charged with electricity of one sign. For the sake of consistency with this fact, Poincaré has assumed a pressure to exist inside these particles which balances the electrostatic repulsion. It cannot, however, be asserted that this pressure vanishes outside the particles. We shall be consistent with this circumstance if, in our phenomenological presentation, we add a pressure term. This must not, however, be confused with a hydrodynamical pressure, as it serves only for the energetic presentation of the dynamical relations inside matter. Accordingly we put

$$T_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} \sigma \frac{dx_\alpha}{ds} \frac{dx_\beta}{ds} - g_{\mu\nu} p \quad . \quad (122)$$

In our special case we have, therefore, to put

$$T_{\mu\nu} = \gamma_{\mu\nu} p \text{ (for } \mu \text{ and } \nu \text{ from 1 to 3)}$$
$$T_{44} = \sigma - p$$
$$T = -\gamma^{\mu\nu}\gamma_{\mu\nu} p + \sigma - p = \sigma - 4p$$

Observing that the field equation (96) may be written in the form

$$R_{\mu\nu} = -\kappa(T_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} T)$$

we get from (96) the equations,

$$+\frac{2}{a^2}\gamma_{\mu\nu} = \kappa\left(\frac{\sigma}{2} - p\right)\gamma_{\mu\nu}$$

$$0 = -\kappa\left(\frac{\sigma}{2} + p\right)$$

From this follows

$$\left.\begin{aligned} p &= -\frac{\sigma}{2} \\ a &= \sqrt{\frac{2}{\kappa\sigma}} \end{aligned}\right\} \quad . \quad . \quad (123)$$

If the universe is quasi-Euclidean, and its radius of curvature therefore infinite, then σ would vanish. But it is improbable that the mean density of matter in the universe is actually zero ; this is our third argument against the assumption that the universe is quasi-Euclidean. Nor does it seem possible that our hypothetical pressure can vanish ; the physical nature of this pressure can be appreciated only after we have a better theoretical knowledge of the electromagnetic field. According to the second of equations (123) the radius, a, of the universe is determined in terms of the total mass, M, of matter, by the equation

$$a = \frac{M\kappa}{4\pi^2} \quad . \quad . \quad . \quad (124)$$

The complete dependence of the geometrical upon the physical properties becomes clearly apparent by means of this equation.

Thus we may present the following arguments against the conception of a space-infinite closed, and for the conception of a space-bounded closed, universe :

1. From the standpoint of the theory of relativity, to postulate a closed universe is very much simpler than

to postulate the corresponding boundary condition at infinity of the quasi-Euclidean structure of the universe.

2. The idea that Mach expressed, that inertia depends upon the mutual action of bodies, is contained, to a first approximation, in the equations of the theory of relativity ; it follows from these equations that inertia depends, at least in part, upon mutual actions between masses. Thereby, Mach's idea gains in probability, as it is an unsatisfactory assumption to make that inertia depends in part upon mutual actions, and in part upon an independent property of space. But this idea of Mach's corresponds only to a finite universe, bounded in space, and not to a quasi-Euclidean, infinite universe. From the standpoint of epistemology it is more satisfying to have the mechanical properties of space completely determined by matter, and this is the case only in a closed universe.

3. An infinite universe is possible only if the mean density of matter in the universe vanishes. Although such an assumption is logically possible, it is less probable than the assumption that there is a finite mean density of matter in the universe.

APPENDIX I

ON THE 'COSMOLOGIC PROBLEM'

SINCE the first edition of this little book some advances have been made in the theory of relativity. Some of these we shall mention here only briefly:

The first step forward is the conclusive demonstration of the existence of the red shift of the spectral lines by the (negative) gravitational potential of the place of origin (see p. 88). This demonstration was made possible by the discovery of so-called 'dwarf stars' whose average density exceeds that of water by a factor of the order 10^4. For such a star (e.g. the faint companion of Sirius), whose mass and radius can be determined,* this red shift was expected, by the theory, to be about twenty times as large as for the sun, and indeed it was demonstrated to be within the expected range.

A second step forward, which will be mentioned briefly, concerns the law of motion of a gravitating body. In the initial formulation of the theory the law of motion for a gravitating particle was introduced as an independent fundamental assumption in addition to the field law of gravitation—see Eq. 90 which asserts that a gravitating particle moves in a geodesic line. This constitutes a hypothetic translation of Galileo's law of inertia to the case of the existence of 'genuine' gravitational fields. It has been shown that this law of motion—generalized to the case of arbitrarily large gravitating masses—can be

* The mass is derived from the reaction on Sirius by spectroscopic means, using the Newtonian laws; the radius is derived from the total lightness and from the intensity of radiation per unit area, which may be derived from the temperature of its radiation.

derived from the field-equations of empty space alone. According to this derivation the law of motion is implied by the condition that the field be singular nowhere outside its generating mass points.

A third step forward, concerning the so-called 'cosmologic problem', will be considered here in detail, in part because of its basic importance, partly also because the discussion of these questions is by no means concluded. I feel urged toward a more exact discussion also by the fact that I cannot escape the impression that in the present treatment of this problem the most important basic points of view are not sufficiently stressed.

The problem can be formulated roughly thus: On account of our observations on fixed stars we are sufficiently convinced that the system of fixed stars does not in the main resemble an island which floats in infinite empty space, and that there does not exist anything like a centre of gravity of the total amount of existing matter. Rather, we feel urged toward the conviction that there exists an average density of matter in space which differs from zero.

Hence the question arises: Can this hypothesis, which is suggested by experience, be reconciled with the general theory of relativity?

First we have to formulate the problem more precisely. Let us consider a finite part of the universe which is large enough so that the average density of matter contained in it is an approximately continuous function of (x_1, x_2, x_3, x_4). Such a subspace can be considered approximately as an inertial system (Minkowski space) to which we relate the motion of the stars. One can arrange it so that the mean velocity of matter relative to this system shall vanish in all directions. There remain the (almost random) motions of the individual stars, similar to the motions of the molecules of a gas. It is essential that the velocities of the stars are known by experience to be very small as compared to the velocity of light. It is therefore feasible for the moment to neglect this relative motion completely, and to consider the stars replaced by material dust without (random) motion of the particles against each other.

The above conditions are by no means sufficient to make the problem a definite one. The simplest and most radical specialization would be the condition: The (naturally measured) density, ρ, of matter is the same everywhere in (four-dimensional) space, the metric is, for a suitable choice of co-ordinates, independent of x_4 and homogeneous and isotropic with respect to x_1, x_2, x_3.

It is this case which I at first considered the most natural idealized description of physical space in the large; it is treated on pages 98–103 of this book. The objection to this solution is that one has to introduce a negative pressure, for which there exists no physical justification. In order to make that solution possible I originally introduced a new member into the equation instead of the above-mentioned pressure, which is permissible from the point of view of relativity. The equations of gravitation thus enlarged were:

$$(R_{ik} - \tfrac{1}{2}g_{ik}R) + \Lambda g_{ik} + \kappa T_{ik} = 0 \quad . \quad (1)$$

where Λ is a universal constant ("cosmologic constant"). The introduction of this second member constitutes a complication of the theory, which seriously reduces its logical simplicity. Its introduction can only be justified by the difficulty produced by the almost unavoidable introduction of a finite average density of matter. We may remark, by the way, that in Newton's theory there exists the same difficulty.

The mathematician Friedman found a way out of this dilemma.* His result then found a surprising confirmation by Hubble's discovery of the expansion of the stellar system (a red shift of the spectral lines which increases uniformly with distance). The following is essentially nothing but an exposition of Friedman's idea:

FOUR-DIMENSIONAL SPACE WHICH IS ISOTROPIC WITH RESPECT TO THREE DIMENSIONS

We observe that the systems of stars, as seen by us, are spaced with approximately the same density in all directions. Thereby we are moved to the assumption that the *spatial* isotropy of the system would hold for all observers, for every place and every time of an observer who is at rest as compared with surrounding matter. On the other hand we no longer make the assumption that the average density of matter, for an observer who is at rest relative to neighbouring matter, is constant with respect to time. With this we drop the assumption that the expression of the metric field is independent of time.

We now have to find a mathematical form for the condition that the universe, *spatially speaking*, is isotropic everywhere. Through every point P of (four-dimensional) space there is the path of a particle (which in the following will be called 'geodesic' for short). Let P and Q be two infinitesimally near points of such a geodesic. We shall then have to demand that the expression of the field shall be invariant relative to any rotation of the co-ordinate system keeping P and Q fixed. This will be valid for any element of any geodesic.†

* He showed that it is possible, according to the field equations, to have a finite density in the whole (three-dimensional) space, without enlarging these field equations *ad hoc*. *Zeitschr. f. Phys.*, 10 (1922).

† This condition not only limits the metric, but it necessitates that for every geodesic there exist a system of co-ordinates such that relative to this system the invariance under rotation around this geodesic is valid.

view of relativity, is to be rejected from the point of view of logical economy. As Friedman was the first to show one can reconcile an everywhere finite density of matter with the original form of the equations of gravity if one admits the time variability of the metric distance of two mass points.*

(2) The demand for *spatial* isotropy of the universe alone leads to Friedman's form. It is therefore undoubtedly the general form, which fits the cosmologic problem.

(3) Neglecting the influence of spatial curvature, one obtains a relation between the mean density and Hubble's expansion which, as to order of magnitude, is confirmed empirically.

One further obtains, for the time from the start of the expansion up to the present, a value of the order of magnitude of 10^9 years. The brevity of this time does not concur with the theories on the developments of fixed stars.

(4) The latter result is not changed by the introduction of spatial curvature; nor is it changed by the consideration of the random motion of stars and systems of stars with respect to each other.

(5) Some try to explain Hubble's shift of spectral lines by means other than the Doppler effect. There is, however, no support for such a conception in the known physical feats. According to such a hypothesis it would be possible to connect two stars, S_1 and S_2, by a rigid rod. Monochromatic light which is sent from S_1 to S_2

* If Hubble's expansion had been discovered at the time of the creation of the general theory of relativity, the cosmologic member would never have been introduced. It seems now so much less justified to introduce such a member into the field equations, since its introduction loses its sole original justification—that of leading to a natural solution of the cosmologic problem.

and reflected back to S_1 could arrive with a different frequency (measured by a clock on S_1) if the number of wave lengths of light along the rod should change with time on the way. This would mean that the locally measured velocity of light would depend on time, which would contradict even the special theory of relativity. Further it should be noted that a light signal going to and fro between S_1 and S_2 would constitute a 'clock' which would not be in a constant relation with a clock (e.g. an atomistic clock) in S_1. This would mean that there would exist no metric in the sense of relativity. This not only involves the loss of comprehension of all those relations which relativity has yielded, but it also fails to concur with the fact that certain atomistic forms are not related by 'similarity' but by 'congruence' (the existence of sharp spectral lines, volumes of atoms, &c.).

The above considerations are, however, based on wave theory, and it may be that some proponents of the above hypothesis imagine that the process of the expansion of light is altogether not according to wave theory, but rather in a manner analogous to the Compton effect. The assumption of such a process without scattering constitutes a hypothesis which is not justified from the point of view of our present knowledge. It also fails to give a reason for the independence of the relative shift of frequence from the original frequency. Hence one cannot but consider Hubble's discovery as an expansion of the system of stars.

(6) The doubts about the assumption of a 'beginning of the world' (start of the expansion) only about 10^9 years ago have roots of both an empirical and a theoretical nature. The astronomers tend to consider the stars of different spectral types as age classes of a uniform development, which process would need much longer than 10^9 years. Such a theory, therefore, actually contradicts

the demonstrated consequences of the relativistic equations. It seems to me, however, that the 'theory of evolution' of the stars rests on weaker foundations than the field equations.

The theoretical doubts are based on the fact that for the time of the beginning of the expansion the metric becomes singular and the density, ρ, becomes infinite. In this connexion the following should be noted: The present theory of relativity is based on a division of physical reality into a metric field (gravitation) on the one hand, and into an electromagnetic field and matter on the other hand. In reality space will probably be of a uniform character and the present theory be valid only as a limiting case. For large densities of field and of matter, the field equations and even the field variables which enter into them will have no real significance. One may not therefore assume the validity of the equations for very high density of field and of matter, and one may not conclude that the 'beginning of the expansion' must mean a singularity in the mathematical sense. All we have to realize is that the equations may not be continued over such regions.

This consideration does, however, not alter the fact that the 'beginning of the world' really constitutes a beginning, from the point of view of the development of the now existing stars and systems of stars, at which those stars and systems of stars did not yet exist as individual entities.

(7) There are, however, some empirical arguments in *favour* of a dynamic concept of space as required by the theory. Why does there still exist uranium, despite its comparatively rapid decomposition, and despite the fact that no possibility for the creation of uranium is recognizable? Why is space not so filled with radiation as to make the nocturnal sky look like a glowing surface? This is an old question which so far has found no satisfactory

GENERAL REMARKS

A. In my opinion the theory presented here is the logically simplest relativistic field theory which is at all possible. But this does not mean that nature might not obey a more complex field theory.

More complex field theories have frequently been proposed. They may be classified according to the following characteristic features:

(*a*) Increase of the number of dimensions of the continuum. In this case one must explain why the continuum is *apparently* restricted to four dimensions.

(*b*) Introduction of fields of different kind (e.g. a vector field) in addition to the displacement field and its correlated tensor field g_{ik} (or g^{ik}).

(*c*) Introduction of field equations of higher order (of differentiation).

In my view, such more complicated systems and their combinations should be considered only if there exist physical-empirical reasons to do so.

B. A field theory is not yet completely determined by the system of field equations. Should one admit the appearance of singularities? Should one postulate boundary conditions? As to the first question, it is my opinion that singularities must be excluded. It does not seem reasonable to me to introduce into a continuum theory points (or lines, etc.) for which the field equations do not hold. Moreover, the introduction of singularities is equivalent to postulating boundary conditions (which are arbitrary from the point of view of the field equations) on 'surfaces' which closely surround the singularities. Without such a postulate the theory is much too vague. In my opinion the answer to the second question is that the postulation of boundary conditions is indispensable. I shall demonstrate this by an elementary example. One can compare the postulation of a potential of the form

$\phi = \sum \frac{m}{r}$ with the statement that outside the mass points (in three dimensions) the equation $\Delta\phi = 0$ is satisfied. But if one does not add the boundary condition that ϕ vanish (or remain finite) at infinity, then there exist solutions that are entire functions of the x (e.g. $x_1^2 - \frac{1}{2}(x_2^2 + x_3^2)$) and become infinite at infinity. Such fields can only be excluded by postulating a boundary condition in case the space is an 'open' one.

C. Is it conceivable that a field theory permits one to understand the atomistic and quantum structure of reality? Almost everybody will answer this question with 'no'. But I believe that at the present time nobody knows anything reliable about it. This is so because we cannot judge in what manner and how strongly the exclusion of singularities reduces the manifold of solutions. We do not possess any method at all to derive systematically solutions that are free of singularities. Approximation methods are of no avail since one never knows whether or not there exists to a particular approximate solution an exact solution *free of singularities*. For this reason we cannot at present compare the content of a nonlinear field theory with experience. Only a significant progress in the mathematical methods can help here. At the present time the opinion prevails that a field theory must first, by 'quantization', be transformed into a statistical theory of field probabilities according to more or less established rules. I see in this method only an attempt to describe relationships of an essentially nonlinear character by linear methods.

D. One can give good reasons why reality cannot at all be represented by a continuous field. From the quantum phenomena it appears to follow with certainty that a finite system of finite energy can be completely described by a finite set of numbers (quantum numbers). This does not seem to be in accordance with a continuum theory, and must lead to an attempt to find a purely algebraic theory for the description of reality. But nobody knows how to obtain the basis of such a theory.