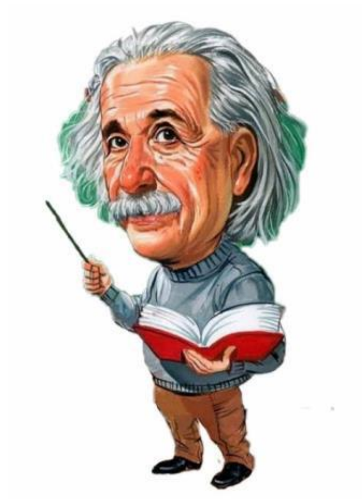
The Scientific Papers Of Albert Einstein

Manjunath R

The Scientific Papers Of Albert Einstein



Born in Germany to Jewish parents, Albert Einstein (14 March 1879 – 18 April 1955) was one of the most celebrated scientists of the Twentieth Century – a well known freethinker, speaking on a range of humanitarian and global issues – who published the special and general theories of relativity and best known for the concept of mass-energy equivalence expressed by the famous equation, E = mc squared. He won the Nobel Prize in physics for his discovery of the law of the photoelectric effect. On 17 April 1955, Einstein experienced internal bleeding caused by the rupture of an abdominal aortic aneurysm, and eventually died of an aortic aneurysm on 18 April 1955 in Princeton Hospital in New Jersey at the age of 76, having continued to search for a "theory of everything" still unrealized.

On the Motion of Small Particles Suspended in Liquids at Rest Required by the Molecular-Kinetic Theory of Heat*

Albert Einstein
*Originally published in *Annalen der Physik* 17 [1905]: 549-560

IN THIS PAPER it will be shown that, according to the molecular-kinetic theory of heat, bodies of a microscopically visible size suspended in liquids must, as a result of thermal molecular motions, perform motions of such magnitude that they can be easily observed with a microscope. It is possible that the motions to be discussed here are identical with so-called Brownian molecular motion; however, the data available to me on the latter are so imprecise that I could not form a judgment on the question. If the motion to be discussed here can actually be observed, together with the laws it is expected to obey, then classical thermodynamics can no longer be viewed as applying to regions that can be distinguished even with a microscope, and an exact determination of actual atomic sizes becomes possible. On the other hand, if the prediction of this motion were to be proved wrong, this fact would provide a far-reaching argument against the molecular-kinetic conception of heat.

1 On the Osmotic Pressure to be Ascribed to Suspended Particles

Let z gram-molecules of a non-electrolyte be dissolved in a part V^* of the total volume V of a liquid. If the volume V^* is separated from the pure solvent by a wall that is permeable to the solvent but not to the solute, then this wall is subjected to a so-called osmotic pressure, which for sufficiently large values of V^*/z satisfies the equation

$$pV^* = RTz. \tag{1}$$

But if instead of the solute, the partial volume V^* of the liquid contains small suspended bodies that also cannot pass through the solvent-permeable wall, then, according to the classical theory of thermodynamics, we should not expect–at least if we neglect the force of gravity, which does not interest us here–any pressure to be exerted on the wall; for according to the usual interpretation, the "free energy" of the system does not seem to depend on the position of the wall and of the suspended bodies, but only on the total mass and properties of the suspended substance, the liquid, and the wall, as well as on the pressure and temperature. To be sure, the energy and entropy of the interfaces (capillary forces) should also be considered when calculating the free energy; but we can disregard them here because changes in the position of the wall and suspended bodies will not cause changes in the size and state of the contact surfaces.

But a different interpretation arises from the standpoint of the molecular-kinetic theory of heat. According to this theory, a dissolved molecule differs from a suspended body only in size, and it is difficult to see why suspended bodies should not produce the same osmotic pressure as an equal number of dissolved molecules. We have to assume that the suspended bodies perform an irregular, albeit very slow, motion in the liquid due to the liquid's molecular motion; if prevented by the wall from leaving the volume V^*, they will exert pressure upon the wall just like molecules in solution. Thus, if n suspended bodies are present in the volume V^*, i.e., $n/V^* = \nu$ in a unit volume, and if neighboring bodies are sufficiently far separated from each other, there will be a corresponding osmotic pressure p of magnitude

$$p = \frac{RT}{V^*}\frac{n}{N} = \frac{RT}{N} \cdot \nu, \tag{2}$$

where N denotes the number of actual molecules per gram-molecule. It will be shown in the next section that the molecular-kinetic theory of heat does indeed lead to this broader interpretation of osmotic pressure.

2 Osmotic Pressure from the Standpoint of the Molecular-Kinetic Theory of Heat [1]

If $p_1 p_2 \dots p_l$ are state variables of a physical system that completely determine the system's instantaneous state (e.g., the coordinates and velocity components of all atoms of the system), and if the complete system of equations for changes of these variables is given in the form

$$\frac{\partial p_\nu}{\partial t} = \varphi_\nu(p_1 \dots p_l) \ \ (\nu = 1, 2, \dots l), \tag{3}$$

with $\sum \frac{\partial \varphi_\nu}{\partial p_\nu} = 0$, then the entropy of the system is given by the expression

$$S = \frac{\bar{E}}{T} + 2\kappa \ln \int e^{-\frac{E}{2\kappa T}} dp_1 \dots dp_l. \tag{4}$$

Here T denotes the absolute temperature, $\bar{E}$ the energy of the system, and E the energy as a function of the p_ν. The integral extends over all possible values of p_ν consistent with the conditions of the problem. κ is connected with the constant N mentioned above by the relation $2\kappa N = R$. Hence we get for the free energy F

$$F = -\frac{R}{N} T \ln \int e^{-\frac{EN}{RT}} dp_1 \dots dp_l = -\frac{RT}{N} \ln B. \tag{5}$$

Let us now imagine a liquid enclosed in volume V; let a part V^* of the volume V contain n solute molecules or suspended bodies, which are retained in the volume V^* by a semipermeable wall; the integration limits of the integral R occurring in the expressions for S and F will be affected accordingly. Let the total volume of the solute molecules or suspended bodies be small compared with V^*. In accordance with the theory mentioned, let this system be completely described by the variables $p_1 \dots p_l$.

Even if the molecular picture were extended to include all details, the calculation of the integral B would he so difficult that an exact calculation of F is hardly conceivable. However, here we only need to know how F depends on the size of the volume V^* in which all the solute molecules or suspended bodies (hereafter called "particles" for brevity) are contained.

Let us call the rectangular coordinates of the center of gravity of the first particle x_1, y_1, z_1, those of the second x_2, y_2, z_2, etc., and those of the last particle x_n, y_n, z_n, and assign to the centers of gravity of the particles the infinitesimally small parallelepiped regions $dx_1 dy_1 dz_1, dx_2 dy_2 dz_2 \dots dx_n dy_n dz_n$, all of which lie in V^*. We want to evaluate the integral occurring in the expression for F, with the restriction that the centers of gravity of the particles shall lie in the regions just assigned to them. In any case, this integral can be put into the form

$$dB = dx_1 dy_1 \dots dz_n \cdot J, \tag{6}$$

[1] In this section it is assumed that the reader is familiar with the author's papers on the foundations of thermodynamics (cf. *Ann. d. Phys.* 9 [1902]: 417 and 11 [1903]: 170). Knowledge of these Papers and of this section of the present paper is not essential for an understanding of the results in the present paper.

where J is independent of dx_1dy_1, etc., as well as of V^*, i.e., of the position of the semipermeable wall. But J is also independent of the particular choice of the *positions* of the center of gravity regions and of the value of V^*, as will be shown immediately. For if a second system of infinitesimally small regions were assigned to the centers of gravity of the particles and denoted by $dx'_1dy'_1dz'_1, dx'_2dy'_2dz'_2 \dots dx'_ndy'_ndz'_n$, and if these regions differed from the originally assigned ones by their position alone, but not by their size, and if, likewise, all of them were contained in V^*, we would similarly have

$$dB' = dx'_1dy'_1 \dots dz'_n \cdot J', \tag{7}$$

where

$$dx_1dy_1 \dots dz_n = dx'_1dy'_1 \dots dz'_n. \tag{8}$$

Hence,

$$\frac{dB}{dB'} = \frac{J}{J'}. \tag{9}$$

But from the molecular theory of heat, as presented in the papers cited,[2] it is easily deduced that dB/B and dB'/B are respectively equal to the probabilities that at an arbitrarily chosen moment the centers of gravity of the particles will be found in the regions $(dx_1 \dots dz_n)$ and $(dx'_l \dots dz'_n)$ respectively. If the motions of the individual particles are independent of one another (to a sufficient approximation) and if the liquid is homogeneous and no forces act on the particles, then for regions of the same size the probabilities of the two systems of regions will be equal, so that

$$\frac{dB}{B} = \frac{dB'}{B}. \tag{10}$$

But from this equation and the previous one it follows that

$$J = J'. \tag{11}$$

This proves that J does not depend on either V^* or $x_1, y_1, \dots z_n$. By integrating, we get

$$B = \int J dx_l \dots dz_n = JV^{*n}, \tag{12}$$

and hence

$$F = -\frac{RT}{N}\{\ln J + n \ln V^*\} \tag{13}$$

and

$$p = -\frac{\partial F}{\partial V^*} = \frac{RT}{V^*}\frac{n}{N} = \frac{RT}{N}\nu. \tag{14}$$

This analysis shows that the existence of osmotic pressure can be deduced from the molecular-kinetic theory of heat, and that, at high dilution, according to this theory, equal numbers of solute molecules and suspended particles behave identically as regards osmotic pressure.

3 Theory of Diffusion of Small Suspended Spheres

Suppose suspended particles are randomly distributed in a liquid. We will investigate their state of dynamic equilibrium under the assumption that a force K, which depends on the position but not on the time, acts on the individual particles. For the sake of simplicity, we shall assume that the force acts everywhere in the direction of the X-axis.

[2] A. Einstein, *Ann. d. Phys.* 11 (1903): 170.

If the number of suspended particles per unit volume is ν, then in the case of thermodynamic equilibrium ν is a function of x such that the variation of the free energy vanishes for an arbitrary virtual displacement δx of the suspended substance. Thus

$$\delta F = \delta E - T\,\delta S = 0. \tag{15}$$

Let us assume that the liquid has a unit cross section perpendicular to the X-axis, and that it is bounded by the planes $x = 0$ and $x = l$. We then have

$$\delta E = -\int_0^l K\nu\,\delta x\,dx \tag{16}$$

and

$$\delta S = \int_0^l R\frac{\nu}{N}\frac{\partial \delta x}{\partial x}dx = -\frac{R}{N}\int_0^l \frac{\partial \nu}{\partial x}\delta x dx. \tag{17}$$

The required equilibrium condition is therefore

$$-K\nu + \frac{RT}{N}\frac{\partial \nu}{\partial x} = 0 \tag{18}$$

or

$$K\nu - \frac{\partial p}{\partial x} = 0. \tag{19}$$

The last equation asserts that the force K is equilibrated by the force of osmotic pressure.

We can use equation (18) to determine the diffusion coefficient of the suspended substance. We can look upon the dynamic equilibrium state considered here as a superposition of two processes proceeding in opposite directions, namely:

1. A motion of the suspended substance under the influence of the force K that acts on each suspended particle.

2. A process of diffusion, which is to be regarded as the result of the disordered motions of the particles produced by thermal molecular motion.

If the suspended particles have spherical form (where P is the radius of the sphere) and the coefficient of viscosity of the liquid is k, then the force K imparts to an individual particle the velocity[3]

$$\frac{K}{6\pi kP}, \tag{20}$$

and

$$\frac{\nu K}{6\pi kP}$$

particles will pass through a unit area per unit time.

Further, if D denotes the diffusion coefficient of the suspended substance and μ the mass of a particle, then

$$-D\frac{\partial(\mu\nu)}{\partial x} \text{ grams} \tag{21}$$

or

$$-D\frac{\partial \nu}{\partial x} \tag{22}$$

[3]Cf., e.g., G. Kirchhoff, *Vorlesungen uber Mechanik*, 26. Vorl., 4 (*Lectures on Mechanics*, Lecture 26, sec. 4).

particles will pass across a unit area per unit time as the result of diffusion. Since dynamic equilibrium prevails, we must have

$$\frac{\nu K}{6\pi kP} - D\frac{\partial \nu}{\partial x} = 0. \tag{23}$$

From the two conditions (18) and (23) found for dvnamic equilibrium, we can calculate the diffusion coefficient. We get

$$D = \frac{RT}{N} \cdot \frac{1}{6\pi kP}. \tag{24}$$

Thus, except for universal constants and the absolute temperature, the diffusion coefficient of the suspended substance depends only on the viscosity of the liquid and on the size of the suspended particles.

4 On the Disordered Motion of Particles Suspended in a Liquid and its Relation to Diffusion

We shall now turn to a closer examination of the disordered motions that arise from thermal molecular motion and give rise to the diffusion investigated in the last section.

Obviously, we must assume that each individual particle executes a motion that is independent of the motions of all the other particles; the motions of the same particle in different time intervals must also be considered as mutually independent processes, so long as we think of these time intervals as chosen not to be too small.

We now introduce a time interval τ, which is very small compared with observable time intervals but still large enough that the notions performed by a particle during two consecutive time intervals τ can be considered as mutually independent events.

Suppose, now, that a total of n suspended particles is present in a liquid. In a time interval τ, the X-coordinates of the individual particles will increase by Δ, where Δ has a different (positive or negative) value for each particle. A certain probability distribution law will hold for Δ: the number dn of particles experiencing a displacement that lies between Δ and $\Delta + d\Delta$ in the time interval τ will be expressed by an equation of the form

$$dn = n\varphi(\Delta)d\Delta, \tag{25}$$

where

$$\int_{-\infty}^{+\infty} \varphi(\Delta)d\Delta = 1, \tag{26}$$

and φ differs from zero only for very small values of Δ and satisfies the condition

$$\varphi(\Delta) = \varphi(-\Delta). \tag{27}$$

We will now investigate how the diffusion coefficient depends on φ, restricting ourselves again to the ease where the number ν of particles per unit volume only depends on x and t.

Let $\nu = f(x,t)$ be the number of particles per unit volume; we calculate the distribution of particles at time $t+\tau$ from their distribution at time t. From the definition of the function $\varphi(\Delta)$ we can easily obtain the number of particles found at time $t+\tau$ between two planes perpendicular to the X-axis with abscissas x and $x+dx$. One obtains

$$f(x,t+\tau)dx = dx \cdot \int_{\Delta=-\infty}^{\Delta=+\infty} f(x+\Delta)\varphi(\Delta)d\Delta. \tag{28}$$

But since τ is very small, we can put

$$f(x,t+\tau) = f(x,t) + \tau \frac{\partial f}{\partial t}. \tag{29}$$

Further, let us expand $f(x+\Delta,t)$ in powers of Δ:

$$f(x+\Delta,t) = f(x,t) + \Delta \frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!} \frac{\partial^2 f(x,t)}{\partial x^2} \ldots \text{ad inf.} \tag{30}$$

We can bring this expansion under the integral sign since only very small values of Δ contribute anything to the latter. We obtain

$$f + \frac{\partial f}{\partial t} \cdot \tau = f \cdot \int_{-\infty}^{+\infty} \varphi(\Delta) d\Delta + \frac{\partial f}{\partial x} \int_{-\infty}^{+\infty} \Delta \varphi(\Delta) d\Delta + \frac{\partial^2 f}{\partial x^2} \int_{-\infty}^{+\infty} \frac{\Delta^2}{2} \varphi(\Delta) d\Delta \ldots \tag{31}$$

On the right-hand side, the second, fourth, etc., terms vanish since $\varphi(x) = \varphi(-x)$, while for the first, third, fifth, etc., terms, each successive term is very small compared with the one preceding it. From this equation, by taking into account that

$$\int_{-\infty}^{+\infty} \varphi(\Delta) d\Delta = 1, \tag{32}$$

and putting

$$\frac{1}{\tau} \int_{-\infty}^{+\infty} \frac{\Delta^2}{2} \varphi(\Delta) d\Delta = D, \tag{33}$$

and taking into account only the first and third terms of the right-hand side, we get

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2}. \tag{34}$$

This is the well-known differential equation for diffusion, and we recognize that D is the diffusion coefficient.

Another important point can be linked to this argument. We have assumed that the individual particles are all referred to the same coordinate system. However, this is not necessary since the motions of the individual particles are mutually independent. We will now refer the motion of each particle to a coordinate system whose origin coincides with the position of the center of gravity of the particle in question at time $t = 0$, with the difference that $f(x,t)dx$ now denotes the number of particles whose X-coordinate has increased between the times $t = 0$ and $t = t$ by a quantity that lies somewhere between x and $x + dx$. Thus, the function f varies according to equation (23) in this case as well. Further, it is obvious that for $x \neq 0$ and $t = 0$ we must have

$$f(x,t) = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} f(x,t) dx = n. \tag{35}$$

The problem, which coincides with the problem of diffusion outwards from a point (neglecting the interaction between the diffusing particles), is now completely determined mathematically; its solution is

$$f(x,t) = \frac{n}{\sqrt{4\pi D}} \frac{e^{-\frac{x^2}{4Dt}}}{\sqrt{t}}. \tag{36}$$

The probability distribution of the resulting displacements during an arbitrary time t is thus the same as the distribution of random errors, which was to be expected. What is important, however, is how the constant in the exponent is related to the diffusion coefficient. With the help of this equation we can now calculate the displacement λ_x in the direction of the X-axis that a particle experiences on the average, or, to be more precise, the root-mean-square displacement in the X-direction; it is

$$\lambda_x = \sqrt{\overline{x^2}} = \sqrt{2Dt}. \tag{37}$$

The mean displacement is thus proportional to the square root of the time. It can easily be shown that the root-mean-square of the *total displacements* of the particles has the value $\lambda_x\sqrt{3}$.

5 Formula for the Mean Displacement of Suspended Particles. A New Method of Determining the Actual Size of Atoms

In section 3 we found the following value for the diffusion coefficient D of a substance suspended in a liquid in the form of small spheres of radius P :

$$D = \frac{RT}{N}\frac{1}{6\pi kP}. \tag{38}$$

Further, we found in section 4 that the mean value of the displacements of the particles in the X-direction at time t equals

$$\lambda_x = \sqrt{2Dt}. \tag{39}$$

By eliminating D, we get:

$$\lambda_x = \sqrt{t}\cdot\sqrt{\frac{RT}{N}\frac{1}{3\pi kP}}. \tag{40}$$

This equation shows how λ_x depends on T, k, and P.

We will now calculate how large λ_x is for one second if N is taken to be $6\cdot 10^{23}$ in accordance with the results of the kinetic theory of gases; water at 17^{o}C ($k = 1.35\cdot 10^{-2}$) is chosen as the liquid, and the diameter of the particles is 0.001 mm. We get

$$\lambda_x = 8\cdot 10^{-5}\ \text{cm} \ = 0.8\ \text{micron}. \tag{41}$$

Therefore, the mean displacement in one minute would be about 6 microns.

Conversely the relation can be used to determine N. We obtain

$$N = \frac{t}{\lambda_x^2}\cdot\frac{RT}{3\pi kP}. \tag{42}$$

Let us hope that a researcher will soon succeed in solving the problem presented here, which is so important for the theory of heat.

DOES THE INERTIA OF A BODY DEPEND UPON ITS ENERGY-CONTENT?

By A. EINSTEIN

September 27, 1905

The results of the previous investigation lead to a very interesting conclusion, which is here to be deduced.

I based that investigation on the Maxwell-Hertz equations for empty space, together with the Maxwellian expression for the electromagnetic energy of space, and in addition the principle that:—

The laws by which the states of physical systems alter are independent of the alternative, to which of two systems of coordinates, in uniform motion of parallel translation relatively to each other, these alterations of state are referred (principle of relativity).

With these principles* as my basis I deduced *inter alia* the following result (§ 8):—

Let a system of plane waves of light, referred to the system of co-ordinates (x, y, z), possess the energy l; let the direction of the ray (the wave-normal) make an angle ϕ with the axis of x of the system. If we introduce a new system of co-ordinates (ξ, η, ζ) moving in uniform parallel translation with respect to the system (x, y, z), and having its origin of co-ordinates in motion along the axis of x with the velocity v, then this quantity of light—measured in the system (ξ, η, ζ)—possesses the energy

$$l^* = l \frac{1 - \frac{v}{c} cos\phi}{\sqrt{1 - v^2/c^2}}$$

where c denotes the velocity of light. We shall make use of this result in what follows.

Let there be a stationary body in the system (x, y, z), and let its energy—referred to the system (x, y, z) be E_0. Let the energy of the body relative to the system (ξ, η, ζ) moving as above with the velocity v, be H_0.

Let this body send out, in a direction making an angle ϕ with the axis of x, plane waves of light, of energy $\frac{1}{2}\mathrm{L}$ measured relatively to (x, y, z), and simultaneously an equal quantity of light in the opposite direction. Meanwhile the body remains at rest with respect to the system (x, y, z). The principle of

*The principle of the constancy of the velocity of light is of course contained in Maxwell's equations.

energy must apply to this process, and in fact (by the principle of relativity) with respect to both systems of co-ordinates. If we call the energy of the body after the emission of light E_1 or H_1 respectively, measured relatively to the system (x, y, z) or (ξ, η, ζ) respectively, then by employing the relation given above we obtain

$$\begin{aligned} E_0 &= E_1 + \frac{1}{2}L + \frac{1}{2}L, \\ H_0 &= H_1 + \frac{1}{2}L\frac{1-\frac{v}{c}cos\phi}{\sqrt{1-v^2/c^2}} + \frac{1}{2}L\frac{1+\frac{v}{c}cos\phi}{\sqrt{1-v^2/c^2}} \\ &= H_1 + \frac{L}{\sqrt{1-v^2/c^2}}. \end{aligned}$$

By subtraction we obtain from these equations

$$H_0 - E_0 - (H_1 - E_1) = L\left\{\frac{1}{\sqrt{1-v^2/c^2}} - 1\right\}.$$

The two differences of the form $H - E$ occurring in this expression have simple physical significations. H and E are energy values of the same body referred to two systems of co-ordinates which are in motion relatively to each other, the body being at rest in one of the two systems (system (x, y, z)). Thus it is clear that the difference $H - E$ can differ from the kinetic energy K of the body, with respect to the other system (ξ, η, ζ), only by an additive constant C, which depends on the choice of the arbitrary additive constants of the energies H and E. Thus we may place

$$\begin{aligned} H_0 - E_0 &= K_0 + C, \\ H_1 - E_1 &= K_1 + C, \end{aligned}$$

since C does not change during the emission of light. So we have

$$K_0 - K_1 = L\left\{\frac{1}{\sqrt{1-v^2/c^2}} - 1\right\}.$$

The kinetic energy of the body with respect to (ξ, η, ζ) diminishes as a result of the emission of light, and the amount of diminution is independent of the properties of the body. Moreover, the difference $K_0 - K_1$, like the kinetic energy of the electron (§ 10), depends on the velocity.

Neglecting magnitudes of fourth and higher orders we may place

$$K_0 - K_1 = \frac{1}{2}\frac{L}{c^2}v^2.$$

From this equation it directly follows that:—

If a body gives off the energy L in the form of radiation, its mass diminishes by L/c^2. The fact that the energy withdrawn from the body becomes energy of radiation evidently makes no difference, so that we are led to the more general conclusion that

The mass of a body is a measure of its energy-content; if the energy changes by L, the mass changes in the same sense by $\mathrm{L}/9 \times 10^{20}$, the energy being measured in ergs, and the mass in grammes.

It is not impossible that with bodies whose energy-content is variable to a high degree (e.g. with radium salts) the theory may be successfully put to the test.

If the theory corresponds to the facts, radiation conveys inertia between the emitting and absorbing bodies.

About this Document

This edition of Einstein's *Does the Inertia of a Body Depend upon its Energy-Content* is based on the English translation of his original 1905 German-language paper (published as *Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig?*, in *Annalen der Physik*. **18**:639, 1905) which appeared in the book *The Principle of Relativity*, published in 1923 by Methuen and Company, Ltd. of London. Most of the papers in that collection are English translations by W. Perrett and G.B. Jeffery from the German *Das Relativatsprinzip*, 4th ed., published by in 1922 by Tuebner. All of these sources are now in the public domain; this document, derived from them, remains in the public domain and may be reproduced in any manner or medium without permission, restriction, attribution, or compensation.

The footnote is as it appeared in the 1923 edition. The 1923 English translation modified the notation used in Einstein's 1905 paper to conform to that in use by the 1920's; for example, c denotes the speed of light, as opposed the V used by Einstein in 1905. In this paper Einstein uses L to denote energy; the italicised sentence in the conclusion may be written as the equation "$m = \mathrm{L}/c^2$" which, using the more modern E instead of L to denote energy, may be trivially rewritten as "$\mathrm{E} = mc^2$".

This edition was prepared by John Walker. The current version of this document is available in a variety of formats from the editor's Web site:

`http://www.fourmilab.ch/`

This new Dover edition, first published in 1956, is an unabridged and unaltered republication of the translation first published in 1926. It is published through special arrangement with Methuen and Co., Ltd., and the estate of Albert Einstein.

Manufactured in the United States of America.

INVESTIGATIONS ON THE THEORY OF THE BROWNIAN MOVEMENT

BY

ALBERT EINSTEIN, PH.D.

EDITED WITH NOTES BY

R. FÜRTH

TRANSLATED BY

A. D. COWPER

WITH 3 DIAGRAMS

DOVER PUBLICATIONS, INC.

INVESTIGATIONS ON THE THEORY OF THE BROWNIAN MOVEMENT

I

ON THE MOVEMENT OF SMALL PARTICLES SUSPENDED IN A STATIONARY LIQUID DEMANDED BY THE MOLECULAR-KINETIC THEORY OF HEAT

IN this paper it will be shown that according to the molecular-kinetic theory of heat, bodies of microscopically-visible size suspended in a liquid will perform movements of such magnitude that they can be easily observed in a microscope, on account of the molecular motions of heat. It is possible that the movements to be discussed here are identical with the so-called " Brownian molecular motion "; however, the information available to me regarding the latter is so lacking in precision, that I can form no judgment in the matter (1).

If the movement discussed here can actually be observed (together with the laws relating to it that one would expect to find), then classical thermodynamics can no longer be looked upon as applicable with precision to bodies even of dimensions distinguishable in a microscope: an exact determination of actual atomic dimensions is #en possible. On the other hand, had the prediction of this movement proved to be incorrect, a weighty argument would be provided against the molecular-kinetic conception of heat.

§ 1. On the Osmotic Pressure to be Ascribed to the Suspended Particles

Let z gram-molecules of a non-electrolyte be dissolved in a volume V^* forming part of a quantity of liquid of total volume V. If the volume V^* is separated from the pure solvent by a partition permeable for the solvent but impermeable for the solute, a so-called "osmotic pressure," p, is exerted on this partition, which satisfies the equation

$$pV^* = RTz \quad . \qquad . \quad (2)$$

when V^*/z is sufficiently great.

On the other hand, if small suspended particles are present in the fractional volume V^* in place of the dissolved substance, which particles are also unable to pass through the partition permeable to the solvent: according to the classical theory of

thermodynamics—at least when the force of gravity (which does not interest us here) is ignored—we would not expect to find any force acting on the partition; for according to ordinary conceptions the "free energy" of the system appears to be independent of the position of the partition and of the suspended particles, but dependent only on the total mass and qualities of the suspended material, the liquid and the partition, and on the pressure and temperature. Actually, for the calculation of the free energy the energy and entropy of the boundary-surface (surface-tension forces) should also be considered; these can be excluded if the size and condition of the surfaces of contact do not alter with the changes in position of the partition and of the suspended particles under consideration.

But a different conception is reached from the standpoint of the molecular-kinetic theory of heat. According to this theory a dissolved molecule is differentiated from a suspended body *solely* by its dimensions, and it is not apparent why a number of suspended particles should not produce the same osmotic pressure as the same number of molecules. We must assume that the suspended particles perform an irregular movement—even if a very slow one—in the liquid, on account of the molecular movement of the liquid; if they are prevented from leaving the volume V^* by the partition, they will exert a pressure on the partition just like molecules in solution. Then, if there are n suspended particles present in the volume V^*, and therefore $n/V^* = \nu$ in a unit of volume, and if neighbouring particles are sufficiently far separated, there will be a corresponding osmotic pressure p of magnitude given by

$$p = \frac{RT}{V^*}\frac{n}{N} = \frac{RT}{N}\cdot\nu,$$

where N signifies the actual number of molecules contained in a gram-molecule. It will be shown in the next paragraph that the molecular-kinetic theory of heat actually leads to this wider conception of osmotic pressure.

§ 2. Osmotic Pressure from the Standpoint of the Molecular-Kinetic Theory of Heat (*)

If $p_1, p_2, \ldots p_l$ are the variables of state of

(*) In this paragraph the papers of the author on the "Foundations of Thermodynamics" are assumed to be familiar to the reader (*Ann. d. Phys.*, **9**, p. 417, 1902; **11**, p. 170, 1903). An understanding of the conclusions reached in the present paper is not dependent on a knowledge of the former papers or of this paragraph of the present paper.

a physical system which completely define the instantaneous condition of the system (for example, the Co-ordinates and velocity components of all atoms of the system), and if the complete system of the equations of change of these variables of state is given in the form

$$\frac{\partial p_\nu}{\partial t} = \phi_\nu(p_1 \ldots p_l)\ (\nu = 1, 2, \ldots l)$$

whence

$$\Sigma\frac{\partial \phi_\nu}{\partial p_\nu} = 0,$$

then the entropy of the system is given by the expression

$$S = \frac{\overline{E}}{T} + 2x\ lg\int e^{-\frac{E}{2xT}} dp_1 \ldots dp_l \quad . \quad (3)$$

where T is the absolute temperature, $\overline{E}$ the energy of the system, E the energy as a function of p_ν. The integral is extended over all possible values of p_ν consistent with the conditions of the problem. x is connected with the constant N referred to before by the relation $2xN = R$. We obtain hence for the free energy F,

$$F = -\frac{R}{N}T\ lg\int e^{-\frac{EN}{RT}} dp_1 \ldots dp_l = -\frac{RT}{N}\ lg\ B.$$

Now let us consider a quantity of liquid enclosed in a volume V; let there be n solute molecules (or suspended particles respectively) in the portion V^* of this volume V, which are retained in the volume V^* by a semi-permeable partition; the integration limits of the integral B obtained in the expressions for S and F will be affected accordingly. The combined. volume of the solute molecules (or suspended particles) is taken as small compared with V^*. This system will be completely defined according to the theory under discussion by the variables of condition $p_1 \ldots p_l$.

If the molecular picture were extended to deal with every single unit, the calculation of the integral B would offer such difficulties that an exact calculation of F could be scarcely contemplated. Accordingly, we need here only to know how F depends on the magnitude of the volume V^*, in which all the solute molecules, or suspended bodies (hereinafter termed briefly "particles") are contained.

We will call x_1, y_1, z_1 the rectangular Co-ordinates of the centre of gravity of the first particle, x_2, y_2, z_2 those of the second, etc., x_n, y_n, z_n those of the last particle, and allocate for the centres of gravity of the particles the indefinitely small domains of parallelopiped form dx_1, dy_1, dz_1; dx_2,

$dy_2, dz_2, \ldots dx_n, dy_n, dz_n,$ lying wholly within V^*. The value of the integral appearing in the expression for F will be sought, with the limitation that the centres of gravity of the particles lie within a domain defined in this manner. The integral can then be brought into the form

$$dB = dx_1\, dy_1 \ldots dz_n \,.\, J,$$

where J is independent of dx_1, dy_1, etc., as well as of V^*, i.e. of the position of the semi-permeable partition. But J is also independent of any special choice of the position of the domains of the centres of gravity and of the magnitude of V^*, as will be shown immediately. For if a second system were given, of indefinitely small domains of the centres of gravity of the particles, and the latter designated $dx_1'dy_1'dz_1'$; $dx_2'dy_2'dz_2'$ $\ldots dx_n'dy_n'dz_n'$, which domains differ from those originally given in their position but not in their magnitude, and are similarly all contained in V^*, an analogous expression holds :—

$$dB' = dx_1'dy_1' \ldots dz_n' \,.\, J'.$$

Whence

$$dx_1dy_1 \ldots dz_n = dx_1'dy_1' \ldots dz_n'.$$

Therefore

$$\frac{dB}{dB'} = \frac{J}{J'}$$

But from the molecular theory of Heat given in the paper quoted,(*) it is easily deduced that dB/B (4) (or dB'/B respectively) is equal to the probability that at any arbitrary moment of time the centres of gravity of the particles are included in the domains $(dx_1 \ldots dz_n)$ or $(dx_1' \ldots dz_n')$ respectively. Now, if the movements of single particles are independent of one another to a sufficient degree of approximation, if the liquid is homogeneous and exerts no force on the particles, then for equal size of domains the probability of each of the two systems will be equal, so that the following holds:

$$\frac{dB}{B} = \frac{dB'}{B'},$$

But from this and the last equation obtained it follows that

$$J = J'.$$

We have thus proved that J is independent both of V^* and of $x_1, y_1, \ldots z_n$. By integration we obtain

$$B = \int J dx_1 \ldots dz_n = J \,.\, V^{*n},$$

and thence

$$F = -\frac{RT}{N}\{lg\ J + n\ lg\ V^*\}$$

(*) A. Einstein, *Ann. d. Phys.*, **11**, p. 170, 1903.

and

$$p = -\frac{\partial F}{\partial V^*} = \frac{RT}{V^*}\frac{n}{N} = \frac{RT}{N}\nu.$$

It has been shown by this analysis that the existence of an osmotic pressure can be deduced from the molecular-kinetic theory of Heat; and that as far as osmotic pressure is concerned, solute molecules and suspended particles are, according to this theory, identical in their behaviour at great dilution.

§ 3. Theory of the Diffusion of Small Spheres in Suspension

Suppose there be suspended particles irregularly dispersed in a liquid. We will consider their state of dynamic equilibrium, on the assumption that a force K acts on the single particles, which force depends on the position, but not on the time. It will be assumed for the sake of simplicity that the force is exerted everywhere in the direction of the x axis.

Let ν be the number of suspended particles per unit volume; then in the condition of dynamic equilibrium ν is such a function of x that the variation of the free energy vanishes for an arbitrary virtual displacement δx of the suspended substance. We have, therefore,

$$\delta F = \delta E - T\delta S = 0.$$

It will be assumed that the liquid has unit area of cross-section perpendicular to the x axis and is bounded by the planes $x = 0$ and $x = l$. We have, then,

$$\delta E = -\int_0^l K\nu\delta x dx$$

and

$$\delta S = \int_0^l R\frac{\nu}{N}\frac{\partial \delta x}{\partial x}dx = -\frac{R}{N}\int_0^l \frac{\partial \nu}{\partial x}\delta x dx.$$

The required condition of equilibrium is therefore

$$-K\nu + \frac{RT}{N}\frac{\partial \nu}{\partial x} = 0 \qquad (1)$$

or

$$K\nu - \frac{\partial p}{\partial x} = 0 \quad . \qquad 5$$

The last equation states that equilibrium with the force K is brought about by osmotic pressure forces.

Equation (1) can be used to find the coefficient of diffusion of the suspended substance. We can look upon the dynamic equilibrium condition considered here as a superposition of two processes proceeding in opposite directions, namely:—

1. A movement of the suspended substance under the influence of the force K acting on each single suspended particle.

2. A process of diffusion, which is to be looked upon as a result of the irregular movement of the particles produced by the thermal molecular movement.

If the suspended particles have spherical form (radius of the sphere = P), and if the liquid has a coefficient of viscosity k, then the force K imparts to the single particles a velocity (*)

$$\frac{K}{6\pi kP} \quad . \quad . \quad . \quad (6)$$

and there will pass a unit area per unit of time

$$\frac{\nu K}{6\pi kP}$$

particles.

If, further, D signifies the coefficient of diffusion of the suspended substance, and μ the mass of a particle, as the result of diffusion there will pass across unit area in a unit of time,

$$-D\frac{\partial(\mu\nu)}{\partial x} \text{ grams}$$

or

$$-D\frac{\partial \nu}{\partial x} \text{ particles.}$$

(*) Cf. e.g. G. Kirchhoff, "Lectures on Mechanics," Lect. 26, § 4.

Since there must be dynamic equilibrium, we must have

$$\frac{\nu K}{6\pi kP} - D\frac{\partial \nu}{\partial x} = 0. \tag{2}$$

We can calculate the coefficient of diffusion from the two conditions (1) and (2) found for the dynamic equilibrium. We get

$$D = \frac{RT}{N}\frac{1}{6\pi kP} \quad . \tag{7}$$

The coefficient of diffusion of the suspended substance therefore depends (except for universal constants and the absolute temperature) only on the coefficient of viscosity of the liquid and on the size of the suspended particles.

§ 4. On the Irregular Movement of Particles Suspended in a Liquid and the Relation of this to Diffusion

We will turn now to a closer consideration of the irregular movements which arise from thermal molecular movement, and give rise to the diffusion investigated in the last paragraph.

Evidently it must be assumed that each single particle executes a movement which is independent of the movement of all other particles; the movements of one and the same particle after

different intervals of time must be considered as mutually independent processes, so long as we think of these intervals of time as being chosen not too small.

We will introduce a time-interval τ in our discussion, which is to be very small compared with the observed interval of time, but, nevertheless, of such a magnitude that the movements executed by a particle in two consecutive intervals of time τ are to be considered as mutually independent phenomena (8).

Suppose there are altogether n suspended particles in a liquid. In an interval of time τ the x-Co-ordinates of the single particles will increase by Δ, where Δ has a different value (positive or negative) for each particle. For the value of Δ a certain probability-law will hold; the number dn of the particles which experience in the time-interval τ a displacement which lies between Δ and $\Delta + d\Delta$, will be expressed by an equation of the form

$$dn = n\phi(\Delta)d\Delta,$$

where

$$\int_{-\infty}^{+\infty}\phi(\Delta)d\Delta = 1$$

and ϕ only differs from zero for very small values of Δ and fulfils the condition

$$\phi(\Delta) = \phi(-\Delta).$$

We will investigate now how the coefficient of diffusion depends on ϕ, confining ourselves again to the case when the number ν of the particles per unit volume is dependent only on x and t.

Putting for the number of particles per unit volume $\nu = f(x, t)$, we will calculate the distribution of the particles at a time $t + \tau$ from the distribution at the time t. From the definition of the function +(*A*), there is easily obtained the number of the particles which are located at the time $t + \tau$ between two planes perpendicular to the x-axis, with abscissæ x and $x + dx$. We get

$$f(x, t+\tau)dx = dx. \int_{\Delta = -m}^{\Delta = +\infty} f(x + \Delta)\phi(\Delta)d\Delta.$$

Now, since τ is very small, we can put

$$f(x, t+\tau) = f(x, t) + \tau\frac{\partial f}{\partial t}.$$

Further, we can expand $f(x + \Delta, t)$ in powers of A :—

$$f(x+\Delta, t) = f(x, t) + \Delta\frac{\partial f(x, t)}{\partial x} + \frac{\Delta^2}{2!}\frac{\partial^2 f(x, t)}{\partial x^2} \ldots \textit{ad inf.}$$

We can bring this expansion under the integral sign, since only very small values of Δ contribute anything to the latter. We obtain

$$f + \frac{\partial f}{\partial t}\cdot\tau = f\int_{-\infty}^{+\infty}\phi(\Delta)d\Delta + \frac{\partial x}{\partial f}\int_{-\infty}^{+\infty}\Delta\phi(\Delta)d\Delta$$
$$+ \frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{+\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta \ldots$$

On the right-hand side the second, fourth, etc., terms vanish since $\phi(x) = \phi(-x)$; whilst of the first, third, fifth, etc., terms, every succeeding term is very small compared with the preceding. Bearing in mind that

$$\int_{-\infty}^{+\infty}\phi(\Delta)d\Delta = 1,$$

and putting

$$\frac{1}{\tau}\int_{-\infty}^{+\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta = D,$$

and taking into consideration only the first and third terms on the right-hand side, we get from this equation

$$\text{(1)} \qquad \frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}.$$

This is the well-known differential equation for diffusion, and we recognise that D is the coefficient of diffusion.

Another important consideration can be related to this method of development. We have assumed that the single particles are all referred to the same Co-ordinate system. But this is unnecessary, since the movements of the single particles are mutually independent. We will now refer the motion of each particle to a co-ordinate system whose origin coincides at the time $t = 0$ with the position of the centre of gravity of the particles in question; with this difference, that $f(x, t)dx$ now gives the number of the particles whose x Co-ordinate has increased between the time $t = 0$ and the time $t = t$, by a quantity which lies between x and $x + dx$. In this case also the function f must satisfy, in its changes, the equation (1). Further, we must evidently have €or $x \gtrless 0$ and $t = 0$,

$$f(x, t) = 0 \text{ and } \int_{-\infty}^{+\infty} f(x, t)dx = n.$$

The problem, which accords with the problem of the diffusion outwards from a point (ignoring possibilities of exchange between the diffusing particles) is now mathematically completely defined (9); the solution is

$$f(x, t) = \frac{n}{\sqrt{4\pi D}} \frac{e^{-\frac{x^2}{4Dt}}}{\sqrt{t}} \quad . \quad . \quad (10)$$

The probable distribution of the resulting displacements in a given time t is therefore the same as that of fortuitous error, which was to be expected. But it is significant how the constants in the exponential term are related to the coefficient of diffusion. We will now calculate with the help

of this equation the displacement λ_x in the direction of the X-axis which a particle experiences on an average, or—more accurately expressed—the square root of the arithmetic mean of the squares of displacements in the direction of the X-axis; it is

$$\lambda_x = \sqrt{\overline{x^2}} = \sqrt{2Dt} \quad . \quad . \quad (11)$$

The mean displacement is therefore proportional to the square root of the time. It can easily be shown that the square root of the mean of the squares of the total displacements of the particles has the value $\lambda_x\sqrt{3}$ . . . (12)

§ 5. Formula for the Mean Displacement of Suspended Particles. A New Method of Determining the Real size of the Atom

In § 3 we found for the coefficient of diffusion D of a material suspended in a liquid in the form of small spheres of radius P—

$$D = \frac{RT}{N} \cdot \frac{1}{6\pi k P}.$$

Further, we found in § 4 for the mean value of the displacement of the particles in the direction of the X-axis in time t—

$$\lambda_x = \sqrt{2Dt}.$$

By eliminating D we obtain

$$\lambda_x = \sqrt{t} \cdot \sqrt{\frac{RT}{N} \frac{1}{3\pi k P}}.$$

This equation shows how λ_x depends on T, k, and P.

We will calculate how great λ_x is for one second, if N is taken equal to $6 \cdot 10^{23}$ in accordance with the kinetic theory of gases, water at 17° C. is chosen as the liquid ($k = 1{\cdot}35 \cdot 10^{-2}$), and the diameter of the particles ·001 mm. We get

$$\lambda_x = 8 \cdot 10^{-5} \text{ cm.} = 0{\cdot}8\mu.$$

The mean displacement in one minute would be, therefore, about 6μ.

On the other hand, the relation found can be used for the determination of N. We obtain

$$N = \frac{1}{\lambda_x^2} \cdot \frac{RT}{3\pi k P}.$$

It is to be hoped that some enquirer may succeed shortly in solving the problem suggested here, which is so important in connection with the theory of Heat. (13)

Berne, *May*, 1905.

(Received, 11 *May*, 1905.)

II

ON THE THEORY OF THE BROWNIAN MOVEMENT

(From the *Annalen der Physik* (4), **19**, 1906, pp. 371-381)

SOON after the appearance of my paper (*) on the movements of particles suspended in liquids demanded by the molecular theory of heat, Siedentopf (of Jena) informed me that he and other physicists—in the first instance, Prof. Gouy (of Lyons)—had been convinced by direct observation that the so-called Brownian motion is caused by the irregular thermal movements of the molecules of the liquid.(†)

Not only the qualitative properties of the Brownian motion, but also the order of magnitude of the paths described by the particles correspond completely with the results of the theory. I will not attempt here a comparison of the slender experimental material at my disposal with the

(*) A. Einstein, *Ann. d. Phys.*, **17**, p. 549, 1905.

(†) M. Gouy, *Journ. de Phys.* (2), **7**, 561, 1888.

III

A NEW DETERMINATION OF MOLECULAR DIMENSIONS

(From the *Annalen der Physik* (4), **19**, 1906, pp. 289-306. Corrections, *ibid.*, **34**, 1911, pp. 591-592.) (23)

THE kinetic theory of gases made possible the earliest determinations of the actual dimensions of the molecules, whilst physical phenomena observable in liquids have not, up to the present, served for the calculation of molecular dimensions. The explanation of this doubtless lies in the difficulties, hitherto unsurpassable, which discourage the development of a molecular kinetic theory of liquids that will extend to details. It will be shown now in this paper that the size of the molecules of the solute in an undissociated dilute solution can be found from the viscosity of the solution and of the pure solvent, and from the rate of diffusion of the solute into the solvent, if the volume of a molecule of the solute is large

compared with the volume of a molecule of the solvent. For such a solute molecule will behave approximately, with respect to its mobility in the solvent, and in respect to its influence on the viscosity of the latter, as a solid body suspended in the solvent, and it will be allowable to apply to the motion of the solvent in the immediate neighbourhood of a molecule the hydrodynamic equations, in which the liquid is considered homogeneous, and, accordingly, its molecular structure is ignored. We will choose for the shape of the solid bodies, which shall represent the solute molecules, the spherical form.

§ 1. On the Effect on the Motion of a Liquid of a Very Small Sphere Suspended in It

As the subject of our discussion, let us take an incompressible homogeneous liquid with viscosity k, whose velocity-components u, v, w will be given as functions of the co-ordinates x, y, z, and of the time. Taking an arbitrary point x_0, y_0, z_0, we will imagine that the functions u, v, w are developed according to Taylor's theorem as functions of $x - x_0$, $y - y_0$, $z - z_0$, and that a domain G is marked out around this point so small that within it only the linear terms in this expansion have to be considered. The motion of the liquid contained in G can then be looked upon in the familiar manner as the result of the superposition of three motions, namely,

1. A parallel displacement of all the particles' of the liquid without change of their relative position.
2. A rotation of the liquid without change of the relative position of the particles of the liquid.
3. A movement of dilatation in three directions at sight angles to one another (the principal axes of dilatation).

We will imagine now a spherical rigid body in the domain G, whose centre lies at the point x_0, y_0, z_0, and whose dimensions are very small compared with those of the domain G, We will further assume that the motion under consideration is so slow that the kinetic energy of the sphere is negligible as well as that of the liquid. It will be further assumed that the velocity components of an element of surface of the sphere show agreement with the corresponding velocity components of the particles of the liquid in the immediate neighbourhood, that is, that the contact-layer (thought of as continuous) also exhibits

everywhere a viscosity-coefficient that is not vanishingly small.

It is clear without further discussion that the sphere simply shares in the partid motions 1 and 2, without modifying the motion of the neighbouring liquid, since the liquid moves as a rigid body in these partial motions; and that we have ignored the effects of inertia.

But the motion 3 will be modified by the presence of the sphere, and our next problem will be to investigate the influence of the sphere on this motion of the liquid. We will further refer the motion 3 to a co-ordinate system whose axes are parallel to the principal axes of dilatation, and we will put

$$x - x_0 = \xi,$$
$$y - y_0 = \eta,$$
$$z - z_0 = \zeta,$$

then the motion can be expressed by the equations

$$(1) \qquad \begin{aligned} u_0 &= A\xi, \\ v_0 &= B\eta, \\ w_0 &= C\zeta, \end{aligned}$$

in the case when the sphere is not present. A, B, C are constants which, on account of the incompressibility of the liquid, must fulfil the condition

$$(2) \qquad A + B + C = 0 \quad . \quad . \quad (24)$$

Now, if the rigid sphere with radius P is introduced at the point x_0, y_0, z_0, the motions of the liquid in its neighbourhood are modified. In the following discussion we will, for the sake of convenience, speak of P as "finite"; whilst the values of ξ, η, ζ, for which the motions of the liquid are no longer appreciably influenced by the sphere, we will speak of as "infinitely great."

Firstly, it is clear from the symmetry of the motions of the liquid under consideration that there can be neither a translation nor a rotation of the sphere accompanying the motion in question, and we obtain the limiting conditions

$$u = v = w = 0 \text{ when } \rho = P$$

where we have put

$$\rho = \sqrt{\xi^2 + \eta^2 + \zeta^2} > 0.$$

Here u, v, w are the velocity-components of the motion now under consideration (modified by the sphere). If we put

$$(3) \qquad \begin{aligned} u &= A\xi + u_1, \\ v &= B\eta + v_1, \\ w &= C\zeta + w_1, \end{aligned}$$

since the motion defined by equation (3) must be transformed into that defined by equations (1) in the "infinite" region, the velocities u_1, v_1, w_1 will vanish in the latter region.

The functions u, v, w must satisfy the hydrodynamic equations with due reference to the viscosity, and ignoring inertia. Accordingly, the following equations will hold :— (*)

$$(4) \left\{ \frac{\partial p}{\partial \xi} = k\Delta u, \ \frac{\partial p}{\partial \eta} = k\Delta v, \ \frac{\partial p}{\partial \zeta} = k\Delta w, \ \frac{\partial u}{\partial \xi} + \frac{\partial v}{\partial \eta} + \frac{\partial w}{\partial \zeta} = 0, \right.$$

where Δ stands for the operator

$$\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} + \frac{\partial^2}{\partial \zeta^2}$$

and p for the hydrostatic pressure.

Since the equations (1) are solutions of the equations (4) and the latter are linear, according to (3) the quantities u_1, v_1, w_1 must also satisfy the equations (4). I have determined u_1, v_1, w_1, and p, according to a method given in the lecture of Kirchhoff quoted in § 4 (†), and find

(*) G. Kirchhoff, "Lectures on Mechanics," Lect. 26.

(†) "From the equations (4) it follows that $\Delta p = 0$. If p is chosen in accordance with this condition, and a function V is determined which satisfies the equation

$$\Delta V = \frac{1}{k} p,$$

then the equations (4) are satisfied if we put

$$u = \frac{\partial V}{\partial \xi} + u', \quad v = \frac{\partial V}{\partial \eta} + v', \quad w = \frac{\partial V}{\partial \zeta} + w'$$

and chose u', v', w', so that $\Delta u' = 0$, $\Delta v' = 0$, and $\Delta w' = 0$, and

$$\frac{\partial u'}{\partial \xi} + \frac{\partial v'}{\partial \eta} + \frac{\partial w'}{\partial \zeta} = -\frac{1}{k} p."$$

$$p = -\frac{5}{3}kP^3\left\{A\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial\xi^2} + B\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial\eta^2}\right.$$

$$\left. + C\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial\zeta^2}\right\} + \text{const.}$$

$$(5)\quad \begin{cases} u = A\xi - \frac{5}{3}P^3 A\frac{\xi}{\rho^3} - \frac{\partial D}{\partial\xi}, \\ v = B\eta - \frac{5}{3}P^3 B\frac{\eta}{\rho^3} - \frac{\partial D}{\partial\eta}, \\ w = C\zeta - \frac{5}{3}P^3 C\frac{\zeta}{\rho^3} - \frac{\partial D}{\partial\zeta}, \end{cases}$$

Now if we put

$$\frac{p}{k} = 2c\frac{\partial^2\frac{1}{\rho}}{\partial\xi^2}$$

and in agreement with this

$$V = c\frac{\partial^2\rho}{\partial\xi^2} + b\frac{\partial^2\frac{1}{\rho}}{\partial\xi^2} + \frac{a}{2}\left(\xi^2 - \frac{\eta^2}{2} - \frac{\zeta^2}{2}\right)$$

and

$$u' = -2c\frac{\partial\frac{1}{\rho}}{\partial\xi},\ v' = 0,\ w' = 0,$$

the constants a, b, c can be chosen so that when $\rho = P$, $u = v = w = 0$. By superposition of three similar solutions we obtain the solution given in the equations (5) and (5a).

where

$$(5a)\quad \begin{cases} D = A\left\{\frac{5}{6}P^3\frac{\partial^2\rho}{\partial\xi^2} + \frac{1}{6}P^5\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial\xi^2}\right\} \\ \quad + B\left\{\frac{5}{6}P^3\frac{\partial^2\rho}{\partial\eta^2} + \frac{1}{6}P^5\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial\eta^2}\right\} \\ \quad + C\left\{\frac{5}{6}P^3\frac{\partial^2\rho}{\partial\zeta^2} + \frac{1}{6}P^5\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial\zeta^2}\right\}. \end{cases}$$

It is easy to see that the equations (5) are solutions of the equations (4). Then, since

$$\Delta\xi = 0,\quad \Delta\frac{1}{\rho} = 0,\quad \Delta\rho = \frac{2}{\rho}$$

and

$$\Delta\left(\frac{\xi}{\rho^3}\right) = -\frac{\partial}{\partial\xi}\left\{\Delta\left(\frac{1}{\rho}\right)\right\} = 0,$$

we get

$$k\Delta u = -k\frac{\partial}{\partial\xi}\{\Delta D\}$$

$$= -k\frac{\partial}{\partial\xi}\left\{\frac{5}{3}P^3 A\frac{\partial^2\frac{1}{\rho}}{\partial\xi^2} + \frac{5}{3}P^3 B\frac{\partial^2\frac{1}{\rho}}{\partial\eta^2} + \dots\right\}.$$

But the last expression obtained is, according to the first of the equations (5), identical with $dp/d\xi$. In similar manner, we can show that the second and third of the equations (4) are satisfied. We obtain further—

$$\frac{\partial u}{\partial \xi} + \frac{\partial v}{\partial \eta} + \frac{\partial w}{\partial \zeta} = (A + B + C)$$

$$+ \frac{5}{3}P^3\left\{A\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial \xi^2} + B\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial \eta^2} + C\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial \zeta^2}\right\} - \Delta D.$$

But since, according to equation (5a),

$$\Delta D = \frac{5}{3}P^3\left\{A\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial \xi^2} + B\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial \eta^2} + C\frac{\partial^2\left(\frac{1}{\rho}\right)}{\partial \zeta^2}\right\},$$

it follows that the last of the equations (4) is satisfied. As for the boundary conditions, our equations for u, v, w are transformed into the equations (1) only when ρ is indefinitely large. By inserting the value of D from the equation (5a) in the second of the equations (5) we get

$$(6) \quad u = A\xi - \frac{5}{2}\frac{P^3}{\rho^5}\xi(A\xi^2 + B\eta^2 + C\zeta^2)$$

$$+ \frac{5}{2}\frac{P^5}{\rho^7}\xi(A\xi^2 + B\eta^2 + C\zeta^2) - \frac{P^5}{\rho^5}A\xi \quad (25)$$

We know that u vanishes when $\rho = P$. On the grounds of symmetry the same holds for v and w. We have now demonstrated that in the equations (5) a solution has been obtained to satisfy both

the equations (4) and the boundary conditions of the problem.

It can also be shown that the equations (5) are the only solutions of the equations (4) consistent with the boundary conditions of the problem. The proof will only be indicated here. Suppose that, in a finite space, the velocity-components of a liquid u, v, w satisfy the equations (4). Now, if another solution U, V, W of the equations (4) can exist, in which on the boundaries of the sphere in question $U = u$, $V = v$, $W = w$, then $(U - u$, $V - v$, $W - w)$ will be a solution of the equations (4), in which the velocity-components vanish at the boundaries of the space. Accordingly, no mechanical work can be done on the liquid contained in the space in question. Since we have ignored the kinetic energy of the liquid, it follows that the work transformed into heat in the space in question is likewise equal to zero. Hence we infer that in the whole space we must have $u = u'$, $v = v'$, $w = w'$, if the space is bounded, at least in part, by stationary walls. By crossing the boundaries, this result can also be extended to the case when the space in question is infinite, as in the case considered above. We can show thus that the solution obtained above is the sole solution of the problem.

We will now place around the point x_0, y_0, z_0 a sphere of radius R, where R is indefinitely large compared with P, and will calculate the energy which is transformed into heat (per unit of time) in the liquid lying within the sphere. This energy W is equal to the mechanical work done on the liquid. If we call the components of the pressure exerted on the surface of the sphere of radius R, X_n, Y_n, Z_n, then

$$W = \int (X_n u + V_n v + Z_n w) ds$$

where the integration is extended over the surface of the sphere of radius R.

Here

$$X_n = -\left(X_\xi \frac{\xi}{\rho} + X_\eta \frac{\eta}{\rho} + X_\zeta \frac{\zeta}{\rho}\right),$$

$$Y_n = -\left(Y_\xi \frac{\xi}{\rho} + Y_\eta \frac{\eta}{\rho} + Y_\zeta \frac{\zeta}{\rho}\right),$$

$$Z_n = -\left(Z_\xi \frac{\xi}{\rho} + Z_\eta \frac{\eta}{\rho} + Z_\zeta \frac{\zeta}{\rho}\right),$$

where

$$X_\xi = p - 2k\frac{\partial u}{\partial \xi}, \qquad Y_\zeta = Z_\eta = -k\left(\frac{\partial v}{\partial \zeta} + \frac{\partial w}{\partial \eta}\right)$$

$$Y_\eta = p - 2k\frac{\partial v}{\partial \eta}, \qquad Z_\xi = X_\zeta = -k\left(\frac{\partial w}{\partial \xi} + \frac{\partial u}{\partial \zeta}\right)$$

$$Z_\zeta = p - 2k\frac{\partial w}{\partial \zeta}, \qquad X_\eta = Y_\xi = -k\left(\frac{\partial u}{\partial \eta} + \frac{\partial v}{\partial \xi}\right).$$

The expressions for u, v, w are simplified when we note that for $\rho = R$ the terms with the factor P^5/ρ^5 vanish.

We have to put

$$\text{(6a)} \qquad \begin{aligned} u &= A\xi - \frac{5}{2}P^3\frac{\xi(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^5} \\ v &= B\eta - \frac{5}{2}P^3\frac{\eta(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^5} \\ w &= C\zeta - \frac{5}{2}P^3\frac{\zeta(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^5} \end{aligned}$$

For p we obtain from the first of the equations (5) by corresponding omissions

$$p = -5kP^3\frac{A\xi^2 + B\eta^2 + C\zeta^2}{\rho^5} + \text{const.}$$

We obtain first

$$\begin{aligned} X_\xi &= -2kA + 10kP^3\frac{A\xi^2}{\rho^5} - 25kP^3\frac{\xi^2(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^7} \\ X_\eta &= +5kP^3\frac{(A+B)\xi\eta}{\rho^5} - 25kP^3\frac{\xi\eta(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^7} \qquad (23) \\ X_\zeta &= +5kP^3\frac{(A+C)\xi\zeta}{\rho^5} - 25kP^3\frac{\xi\zeta(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^7} \end{aligned}$$

and from this

$$X_n = 2Ak\frac{\xi}{\rho} - 5AkP^3\frac{\xi}{\rho^4} + 20kP^3\frac{\xi(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^6}. \qquad (23)$$

With the aid of the expressions for Y_n and Z_n, obtained by cyclic exchange, we get, ignoring all terms which involve the ratio P/ρ raised to any power higher than the third,

$$X_n u + Y_n v + Z_n w = \frac{2k}{\rho}(A^2\xi^2 + B^2\eta^2 + C^2\zeta^2)$$
$$-5k\frac{P^3}{\rho^4}(A^2\xi^2 + B^2\eta^2 + C^2\zeta^2) + 15k\frac{P^3}{\rho^6}(A\xi^2 + B\eta^2 + C\zeta^2)^2. \quad (23)$$

If we integrate over the sphere and bear in mind that

$$\int ds = 4R^2\pi,$$
$$\int \xi^2 ds = \int \eta^2 ds = \int \zeta^2 ds = \tfrac{4}{3}\pi R^4,$$
$$\int \xi^4 ds = \int \eta^4 ds = \int \zeta^4 ds = \tfrac{4}{5}\pi R^6,$$
$$\int \eta^2\zeta^2 ds = \int \zeta^2\xi^2 ds = \int \xi^2\eta^2 ds = \tfrac{4}{15}\pi R^6,$$
$$\int (A\xi^2 + B\eta^2 + C\zeta^2)^2 ds = \tfrac{8}{15}\pi R^6(A^2 + B^2 + C^2), \quad (23)$$

we obtain

$$(7) \quad W = \tfrac{8}{3}\pi R^3 k\delta^2 + \tfrac{4}{3}\pi P^3 k\delta^2 = 2\delta^2 k\left(V + \frac{\Phi}{2}\right), \quad (23)$$

where we put

$$\delta^2 = A^2 + B^2 + C^2,$$
$$\frac{4\pi}{3}R^3 = V \text{ and } \frac{4}{3}\pi P^3 = \Phi$$

If the suspended sphere were not present ($\Phi = 0$), then we should get for the energy used up in the volume V,

$$(7a) \qquad W = 2\delta^2 k V.$$

On account of the presence of the sphere, the energy used up is therefore diminished by $\delta^2 k\Phi$. (26)

§ 2. Calculation of the Viscosity-Coefficient op a Liquid in which a Large Number of Small Spheres are Suspended in Irregular Distribution

In the preceding discussion we have considered the case when there is suspended in a domain G, of the order of magnitude defined above, a sphere that is very small compared with this domain, and have investigated how this influenced the motion of the liquid. We will now assume that an indefinitely large number of spheres are distributed in the domain G, of similar radius and actually so small that the volume of all the spheres together is very small compared with the domain G. Let the number of spheres present in unit volume be n, where n is sensibly constant everywhere in the liquid.

We will now start once more from the motion of a homogeneous liquid, without suspended spheres, and consider again the most general motion of dilatation. If no spheres are present, by suitable choice of the co-ordinate system we can express the velocity components u_0, v_0, w_0, in the arbitrarily-chosen point x, y, z in the domain G, by the equations

$$u_0 = Ax,$$
$$v_0 = By,$$
$$w_0 = Cz,$$

where $$A + B + C = 0.$$

Now a sphere suspended at the point x_ν, y_ν, z_ν, will affect this motion in a manner evident from the equation (6). Since we have assumed that the average distance between neighbouring spheres is very great compared with their radius, and consequently the additional velocity-components originating from all the suspended spheres together are very small compared with u_0, v_0, w_0, we get for the velocity-components u, v, w in the liquid, taking into account the suspended spheres and neglecting terms of higher orders—

$$(8)\quad \begin{cases} u = Ax - \Sigma\left\{\dfrac{5}{2}\dfrac{P^3}{\rho_\nu^2}\dfrac{\xi_\nu(A\xi_\nu^2 + B\eta_\nu^2 + C\zeta_\nu^2)}{\rho_\nu^3} \right. \\ \qquad \left. - \dfrac{5}{2}\dfrac{P^3}{\rho_\nu^4}\dfrac{\xi_\nu(A\xi_\nu^2 + B\eta_\nu^2 + C\zeta_\nu^2)}{\rho_\nu^3} + \dfrac{P^5}{\rho_\nu^4}\dfrac{A\xi_\nu}{\rho_\nu}\right\}, \\ v = By - \Sigma\left\{\dfrac{5}{2}\dfrac{P^3}{\rho_\nu^2}\dfrac{\eta_\nu(A\xi_\nu^2 + B\eta_\nu^2 + C\zeta_\nu^2)}{\rho_\nu^3} \right. \\ \qquad \left. - \dfrac{5}{2}\dfrac{P^5}{\rho_\nu^4}\dfrac{\eta_\nu(A\xi_\nu^2 + B\eta_\nu^2 + C\zeta_\nu^2)}{\rho_\nu^3} + \dfrac{P^5}{\rho_\nu^4}\dfrac{B\eta_\nu}{\rho_\nu}\right\}, \\ w = Cz - \Sigma\left\{\dfrac{5}{2}\dfrac{P^3}{\rho_\nu^2}\dfrac{\zeta_\nu(A\xi_\nu^2 + B\eta_\nu^2 + C\zeta_\nu^2)}{\rho_\nu^3} \right. \\ \qquad \left. - \dfrac{5}{2}\dfrac{P^5}{\rho_\nu^4}\dfrac{\zeta_\nu(A\xi_\nu^2 + B\eta_\nu^2 + C\zeta_\nu^2)}{\rho_\nu^3} + \dfrac{P^5}{\rho_\nu^4}\dfrac{C\zeta_\nu}{\rho_\nu}\right\}, \end{cases}$$

where the summation is extended over ail spheres in the domain G, and we put

$$\xi_\nu = x - x_\nu,$$
$$\eta_\nu = y - y_\nu, \qquad \rho_\nu = \sqrt{\xi_\nu^2 + \eta_\nu^2 + \zeta_\nu^2},$$
$$\zeta_\nu = z - z_\nu.$$

x_ν, y_ν, z_ν are the Co-ordinates of the centre of the sphere.. Further, we 'conclude from the equations (7) and (7a) that the presence of each of the spheres has a result (neglecting indefinitely small quantities of a higher order) (23) in an increase of the heat production per unit volume, and that the energy per unit volume transformed into heat in the domain G has the value

$$W = 2\delta^2 k + n\delta^2 k\Phi, \quad . \quad . \quad (23)$$

or

$$(7b) \qquad W = 2\delta^2 k\left(1 + \frac{\phi}{2}\right), \quad . \quad . \quad (23)$$

where ϕ denotes the fraction of the volume occupied by the spheres.

From the equation (7b) the viscosity-coefficient can be calculated of the heterogeneous mixture of liquid and suspended spheres (hereafter termed briefly "mixture") under discussion; but we must bear in mind that A, B, C are not the values of the principal dilatations in the motion of the liquid defined by the equations (8), (23); we will call the principal dilatations of the mixture A^*, B^*, C^*. On the grounds of symmetry it follows that the principal directions of dilatation of the mixture are parallel to the directions of the principal dilatations A, B, C, and therefore to the Co-ordinate axes. If we write the equations (8) in the form

$$u = Ax + \Sigma u_\nu,$$
$$v = By + \Sigma v_\nu,$$
$$w = Cz + \Sigma w_\nu,$$

we get

$$A^* = \left(\frac{\partial u}{\partial x}\right)_{x=0} = A + \Sigma\left(\frac{\partial u_\nu}{\partial x}\right)_{x=0} = A - \Sigma\left(\frac{\partial u_\nu}{\partial x_\nu}\right)_{x=0}.$$

If we exclude from our discussion the immediate neighbourhood of the single spheres, we can omit the second and third terms of the expressions for $u, v, w,$ and obtain when $x = y = z = 0$:—

$$(9) \quad \begin{cases} u_\nu = -\frac{5}{2}\frac{P^3}{r_\nu^2}\frac{x_\nu(Ax_\nu^2 + By_\nu^2 + Cz_\nu^2)}{r_\nu^3}, \\ v_\nu = -\frac{5}{2}\frac{P^3}{r_\nu^2}\frac{y_\nu(Ax_\nu^2 + By_\nu^2 + Cz_\nu^2)}{r_\nu^3}, \\ w_\nu = -\frac{5}{4}\frac{P^3}{r_\nu^2}\frac{z_\nu(Ax_\nu^2 + Bz_\nu^2 + Gz_\nu^2)}{r_\nu^3} \end{cases}$$

where we put

$$r_\nu = \sqrt{x_\nu^2 + y_\nu^2 + z_\nu^2} > 0.$$

We extend the summation throughout the volume of a sphere K of very large radius R, whose centre lies at the origin of the Co-ordinate system. If we assume further that the irregularly distributed spheres are now evenly distributed and introduce an integral in place of the summation, we obtain

$$A^* = A - n\int_K \frac{\partial u_\nu}{\partial x_\nu} dx_\nu dy_\nu dz_\nu,$$
$$= A - n\int \frac{u_\nu x_\nu}{r_\nu} ds \quad . \quad . \quad (27)$$

where the last integration is to be extended over the surface of the sphere K. Having regard to (9) we find

$$A^* = A - \frac{5}{2}\frac{P^3}{R^6}n \int x_0^2(Ax_0^2 + By_0^2 + Cz_0^2)ds$$
$$= A - n\left(\frac{4}{3}P^3\pi\right)A = A(1 - \phi).$$

By analogy

$$B^* = B(1 - \phi),$$
$$C^* = C(1 - \phi).$$

We will put.

$$\delta^{*2} = A^{*2} + B^{*2} + C^{*2},$$

then neglecting indefinitely small quantities of higher order,

$$\delta^{*2} = \delta^2(1 - 2\phi).$$

We have found for the development of heat per unit of time and volume

$$W^* = 2\delta^2 k\left(1 + \frac{\phi}{2}\right) \quad . \quad . \quad (23)$$

Let us call the viscosity-coefficient of the mixture k^*, then

$$W^* = 2\delta^{*2} k^*.$$

From the last three equations we obtain (neglecting indefinitely small quantities of higher order)

$$k^* = k(1 + 2 \cdot 5\phi) \quad . \quad . \quad (23)$$

We reach, therefore, the result :—

If very small rigid spheres are suspended in a liquid, the coefficient of internal friction is thereby increased by a fraction which is equal to 2·5 times the total volume of the spheres suspended in a unit volume, provided that this total volume is very small.

§ 3. On the Volume of a Dissolved Substance of Molecular Volume Large in Comparison with that of the Solvent

Consider a dilute solution of a substance which does not dissociate in the solution. Suppose that a molecule of the dissolved substance is large compared with a molecule of the solvent; and can be thought of as a rigid sphere of radius P. We can then apply the result obtained in Paragraph 2.

If k^* be the viscosity of the solution, k that of the pure solvent, then

$$\frac{k^*}{k} = 1 + 2 \cdot 5\phi, \quad . \quad . \quad . \quad (23)$$

where ϕ is the total volume of the molecules present in the solution per unit volume.

We will calculate ϕ for a 1 per cent. aqueous sugar solution. According to the observations of Burkhard (Landolt and Börnstein Tables) $k^*/k = 1 \cdot 0245$ (at 20° C.) for a 1 per cent. aqueous sugar solution; therefore $\phi = 0 \cdot 0245$ for (approximately) 0·01 gm. of sugar. A gram of sugar dissolved in water has therefore the same effect on the viscosity as small suspended rigid spheres of total volume 0·98 c.c. (23)

We must recollect here that 1 gm. of solid sugar has the volume 0·61 c.c. We shall find the same value for the specific volume s of the sugar present in solution if the sugar solution is looked upon as a mixture of water and sugar in a dissolved form. The specific gravity of a 1 per cent. aqueous sugar solution (referred to water at the same temperature) at 17·5° is 1·00388. We have then (neglecting the difference in the density of water at 4° and at 17·5°)—

$$\frac{1}{1 \cdot 00388} = 0 \cdot 99 + 0 \cdot 01 s.$$

Therefore $$s = 0 \cdot 61.$$

While, therefore, the sugar solution behaves, as to its density, like a mixture of water and solid sugar, the effect on the viscosity is one and one-half times greater than would have resulted from the suspension of an equal mass of sugar. It appears to me that this result can hardly be explained in the light of the molecular theory, in any other manner than by assuming that the sugar molecules present in solution limit the mobility of the water immediately adjacent, so that a quantity of water, whose volume is approximately one-half (23) the volume of the sugar-molecule, is bound on to the sugar-molecule.

We can say, therefore, that a dissolved sugar molecule (or the molecule together with the water held bound by it respectively) behaves in hydrodynamic relations as a sphere of volume $0{\cdot}98 \cdot 342/N$ c.c. (23), where 342 is the molecular weight of sugar and N the number of actual molecules in a gram-molecule.

§ 4. On the Diffusion of an Undissociated Substance in Solution in a Liquid

Consider such a solution as was dealt with in Paragraph 3. If a force K acts on the molecule, which we will imagine as a sphere of radius P, the molecule will move with a velocity ω which

is determined by P and the viscosity k of the solvent.

That is, the equation holds :—(*)

(1) $$\omega = \frac{k}{6\pi k P} \quad . \quad . \quad . \quad (6)$$

K - force

We will use this relation for the calculation of the diffusion-coefficient of an undissociated solution. If p is the osmotic pressure of the dissolved substance, which is looked upon as the only force producing motion in the dilute solution under consideration, then the force exerted in the direction of the X-axis on the dissolved substance per unit volume of the solution $= -dp/dx$. If there are ρ grams in a unit volume and m is the molecular weight of the dissolved substance, N the number of actual molecules' in a gram-molecule, then $(\rho/m)N$ is the number of (actual) molecules in a unit of volume, and the force acting on a molecule as a result of the fall in concentration will be

(2) $$K = -\frac{m}{\rho N}\frac{\partial p}{\partial x}.$$

If the solution is sufficiently dilute, the osmotic pressure is given by the equation

(3) $$p = \frac{R}{m}\rho T,$$

(*) G. Kirchhoff, "Lectures on Mechanics," Lect. 26 (22).

where T is the absolute temperature and $R = 8\cdot 31 \cdot 10^7$. From the equations (1), (2), and (3) we obtain for the velocity of movement of the dissolved substance

$$\omega = -\frac{RT}{6\pi k}\frac{1}{NP}\frac{\partial\rho}{\partial x}.$$

Finally, the weight of substance passing per unit of time across unit area in the direction of the X-axis will be

$$\omega\rho = -\frac{RT}{6\pi k}\cdot\frac{1}{NP}\frac{\partial\rho}{\partial x}. \tag{4}$$

We, obtain therefore for the diffusion coefficient D—

$$D = \frac{RT}{6\pi k}\cdot\frac{1}{NP}.$$

Accordingly, we can calculate from the diffusion-coefficient and the coefficient of viscosity of the solvent, the value of the product of the number N of actual molecules in a gram-molecule and of the hydrodynamically-effective radius P of the molecule.

In this calculation osmotic pressure is treated as a force acting on the individual molecules, which evidently does not correspond with the conceptions of the kinetic-molecular theory, since, according to the latter, the osmotic pressure in

the case under discussion must be thought of as a virtual force only. However, this difficulty vanishes if we reflect that (dynamic) equilibrium with the (virtual) osmotic forces, which correspond to the differences in concentration of the solution, can be established by the aid of a numerically equal force acting on the single molecules in the opposite direction; as can easily be established following thermodynamic methods.

Equilibrium can be obtained with the osmotic force acting on unit mass, $-\frac{1}{\rho}\frac{\partial\rho}{\partial x}$, by the force $-Px$ (applied to the individual solute molecules) if

$$-\frac{1}{\rho}\frac{\partial\rho}{\partial x} - Px = 0.$$

If we imagine, therefore, two mutually eliminating systems of forces Px and $-Px$ applied to the dissolved substance (per unit mass), then $-Px$ establishes equilibrium with the osmotic pressure and only the force Px, numerically equal to the osmotic pressure, remains over as cause of motion. Thus the difficulty mentioned is overcome.(*)

(*) A detailed statement of this train of thought will be found in *Ann. d. Phys.*, **17**, 1905, p. 549.

§ 5. Determination of Molecular Dimensions with the Help of the Relations already Obtained

We found in Paragraph 3

$$\frac{k^*}{k} = 1 + 2{\cdot}5\phi = 1 + 2{\cdot}5n \, . \, \frac{4}{3}\pi P^3 \qquad (23)$$

where n is the number of solute molecules per unit volume and P the hydrodynamically-effective radius of the molecule. If we bear in mind that

$$\frac{N}{n} = \frac{\rho}{m}$$

where ρ is the mass of the dissolved substance present in unit volume and m is its molecular weight, we obtain

$$NP^3 = \frac{3}{10\pi}\frac{m}{\rho}\left(\frac{k^*}{k} - 1\right).$$

On the other hand, we found in § 4

$$NP = \frac{RT}{6\pi k}\frac{1}{D},$$

These two equations put us in the position to calculate each of the quantities P and N, of which N must show itself to be independent of the nature of the solvent, of the solute and of the temperature, if our theory is to correspond with the facts.

We will carry out the calculation for an aqueous sugar solution. Firstly, it follows from the data given above for the viscosity of sugar solution at 20° C.

$$NP^3 = 80 \quad . \quad . \quad . \quad (23)$$

According to the researches of Graham (calculated out by Stephan), the diffusion-coefficient of sugar in water at 9·5° is 0·384, if the day is taken as unit of time. The viscosity of water at 9·5° is 0·0135. We will insert these data in our formula for the diffusion-coefficient, although they were obtained with 10 per cent. solutions, and it is not to be expected that our formula will be precisely valid at so high a concentration. We get

$$NP = 2{\cdot}08 \, . \, 10^{16}.$$

It follows from the values found for NP^3 and NP, if we ignore the difference in P at 9·5° and 20°, that

$$P = 6{\cdot}2 \, . \, 10^{-8} \text{ cm.} \quad . \quad . \quad (23)$$

$$N = 3{\cdot}3 \, . \, 10^{23}.$$

The value found for N agrees satisfactorily, in order of magnitude, with the values obtained by other methods for this quantity.

Berne, 30 *April*, 1905.

(Received, 19 *August*, 1905.)

Supplement

In the new edition of Landolt and Börnstein's "Physical-Chemical Tables" will be found very useful data for the calculation of the size of the sugar molecule, and the number N of the actual molecules in a gram-molecule. Thovert found (Table, p. 372) for the diffusion-coefficient of sugar in water at 18·5° C. and the concentration 0·005 mol./litre the value 0·33 cm.²/day. From a table (p. 81), with the results of observations made by Hosking, we find by interpolation that in dilute sugar solutions an increase in the sugar-content of 1 per cent. at 18·5° C. corresponds to an increase of the viscosity of 0·00025. Utilizing these data, we find

$$P = 0{\cdot}49 \,.\, 10^{-6} \text{ mm.}$$

and

$$N = 6{\cdot}56 \,.\, 10^{23}. \qquad (23),\ (28)$$

Berne, *January*, 1906.

On a Heuristic Point of View about the Creation and Conversion of Light

On a Heuristic Point of View about the Creation and Conversion of Light (1905)
by Albert Einstein, translated by Wikisource

Maxwell's theory of electromagnetic processes in so-called empty space differs in a profound, essential way from the current theoretical models of gases and other matter. On the one hand, we consider the state of a material body to be determined completely by the positions and velocities of a finite number of atoms and electrons, albeit a very large number. By contrast, the electromagnetic state of a region of space is described by continuous functions and, hence, cannot be determined exactly by any finite number of variables. Thus, according to Maxwell's theory, the energy of purely electromagnetic phenomena (such as light) should be represented by a continuous function of space. By contrast, the energy of a material body should be represented by a discrete sum over the atoms and electrons; hence, the energy of a material body cannot be divided into arbitrarily many, arbitrarily small components. However, according to Maxwell's theory (or, indeed, any wave theory), the energy of a light wave emitted from a point source is distributed continuously over an ever larger volume.

The wave theory of light with its continuous spatial functions has proven to be an excellent model of purely optical phenomena and presumably will never be replaced by another theory. Nevertheless, we should consider that optical experiments observe only time-averaged values, rather than instantaneous values. Hence, despite the perfect agreement of Maxwell's theory with experiment, the use of continuous spatial functions to describe light may lead to contradictions with experiments, especially when applied to the generation and transformation of light.

In particular, black body radiation, photoluminescence, generation of cathode rays from ultraviolet light and other phenomena associated with the generation and transformation of light seem better modeled by assuming that the energy of light is distributed discontinuously in space. According to this picture, the energy of a light wave emitted from a point source is *not* spread continuously over ever larger volumes, but consists of a finite number of energy quanta that are spatially localized at points of space, move without dividing and are absorbed or generated only as a whole.

Subsequently, I wish to explain the reasoning and supporting evidence that led me to this picture of light, in the hope that some researchers may find it useful for their experiments.

Contents

A certain problem concerning the theory of "black body radiation".

We begin by applying Maxwell's theory of light and electrons to the following situation. Let there be a cavity with perfectly reflecting walls, filled with a number of freely moving electrons and gas molecules that interact via conservative forces whenever they come close, i.e., that collide with each other just as gas molecules in the kinetic theory of gases.[1] In addition, let there be a number of electrons bound to spatially well-separated points by restoring forces that increase linearly with separation. These electrons also interact with the free molecules and electrons by conservative potentials when they approach very closely. We denote these electrons, which are bound at points of space, as "resonators", since they absorb and emit electromagnetic waves of a particular period.

According to the present theory of the generation of light, the radiation in the cavity must be identical to black body radiation (which may be found by assuming Maxwell's theory and dynamic equilibrium), at least if one assumes that resonators exist for every frequency under consideration.

Initially, let us neglect the radiation absorbed and emitted by the resonators and focus instead on the requirement of thermal equilibrium and its implications for the interaction (collisions) between molecules and electrons. According to the kinetic theory of gases, dynamic equilibrium requires that the average kinetic energy of a resonator equal the average kinetic energy of a freely moving gas molecule. Decomposing the motion of a resonator electron into three mutually perpendicular oscillations, we find that the average energy $\bar{E}$ of such a linear oscillation is

$$\bar{E} = \frac{R}{N}T,$$

where R is the absolute gas constant, N is the number of "real molecules" in a gram equivalent and T is the absolute temperature. Because of the time averages of the kinetic and potential energy, the energy $\bar{E}$ is ⅔ as large as the kinetic energy of a single free gas molecule. Even if something (such as radiative processes) causes the time-averaged energy of a resonator to deviate from the value $\bar{E}$, collisions with the free electrons and gas molecules will return its average energy to $\bar{E}$ by absorbing or releasing energy. Hence, in this situation, dynamic equilibrium can only exist when every resonator has an average energy $\bar{E}$.

We apply a similar consideration now to the interaction between the resonators and the ambient radiation within the cavity. For this case, Planck has derived the necessary condition for dynamic equilibrium [2]; treating the radiation as a completely random process.[3]

He found:

$$\bar{E}_\nu = \frac{L^3}{8\pi\nu^2}\rho_\nu.$$

Here, $\bar{E}_\nu$ is the average energy of a resonator of eigenfrequency ν (per oscillatory component), L is the speed of light, ν is the frequency, and $\varrho_\nu d\nu$ is the energy density of the cavity radiation of frequency between ν and $\nu + d\nu$.

If the net radiative energy of frequency ν is not to continually increase or decrease, the following equality must hold

$$\frac{R}{N}T = \bar{E} = \bar{E}_\nu = \frac{L^3}{8\pi\nu^2}\rho_\nu,$$

or, equivalently,

$$\rho_\nu = \frac{R}{N}\frac{8\pi\nu^2}{L^3}T.$$

This condition for dynamic equilibrium not only lacks agreement with experiment, it also eliminates any possibility for equilibrium between matter and aether. The wider the range of frequencies of the resonators is chosen the bigger the radiation energy in the space becomes, and in the limit we obtain:

$$\int_0^\infty \rho_\nu \, d\nu = \frac{R}{N}\frac{8\pi}{L^3}T\int_0^\infty \nu^2 \, d\nu = \infty \ .$$

Planck's Derivation of the Fundamental Quantum

In the next section we want to show that the determination that Mr. Planck gave of the elementary quanta is to some extent independent of the "black body radiation" theory that he created.

The Formula by Planck [4] for ϱ_ν that suffices for all experiments so far goes

$$\rho_\nu = \frac{\alpha\nu^3}{e^{\frac{\beta\nu}{T}} - 1},$$

where

$$\alpha = 6.1 \cdot 10^{-56},$$
$$\beta = 4.866 \cdot 10^{-11}.$$

In the limit of large values of T/ν, that is for large wavelengths and radiation densities this formula approaches the form:

$$\rho_\nu = \frac{\alpha}{\beta}\nu^2 T.$$

One recognizes that this formula is the same as the one that was derived from Maxwell theory and electron theory. Equating the coefficients of the formula's:

$$\frac{R}{N}\frac{8\pi}{L^3} = \frac{\alpha}{\beta}$$

or

$$N = \frac{\beta}{\alpha}\frac{8\pi R}{L^3} = 6.17 \cdot 10^{23},$$

that is, a hydrogen atom weighs $1/N$ gram = $1.62 \cdot 10^{-24}$g. This is precisely the value found by Mr. Planck, which is in satisfactory agreement with values obtained in other ways.

This brings us to the conclusion: the larger the energy density and the wavelength of radiation the more suitable the theoretical basis that we used; but for small wavelengths and low radiation densities the basis fails completely.

In the following the "black body radiation" is to be considered in terms of what is experienced, without forming a picture of the creation and propagation of the radiation.

The Entropy of Radiation

The following discussion is contained in a famous work of Mr. Wien, and is only included here for the sake of completeness.

Let there be radiation taking up volume v. We assume that the observable properties of the radiation are determined completely when the radiation densities $\varrho(\nu)$ are given for all frequencies. [5] Since we can regard radiations of different frequency as separable without doing work or transferring heat the entropy of the radiation can be expressed in the form

$$S = v \int_0^\infty \phi(\rho, \nu) d\nu$$

where ϕ is a function of the variables ϱ and ν. ϕ can be reduced to a function of only one variable by expressing that the entropy of radiation between reflecting walls is not changed by adiabatic compression. We won't go into that however, but investigate right away how the function ϕ can be obtained from the radiation law of the black body.

In the case of "black body radiation" ϱ is such a function of ν that for a given energy the entropy is a maximum, that is, that

$$\delta \int_0^\infty \phi(\rho, \nu) d\nu = 0,$$

When

$$\delta \int_0^\infty \rho d\nu = 0.$$

From this it follows that for any choice of δϱ as function of ν

$$\int_0^\infty \left(\frac{\partial \phi}{\partial \rho} - \lambda \right) \delta \rho d\nu = 0,$$

Where λ is independent of ν. Thus $\partial\phi/\partial\rho$ is independent of ν

For the temperature increase of *dT* of a black body radiation of volume *v* = 1 the following equation is valid:

$$dS = \int_{\nu=0}^{\nu=\infty} \frac{\partial \phi}{\partial \rho} d\rho d\nu,$$

or, since $\partial\phi/\partial\rho$ is independent of ν:

$$dS = \frac{\partial \phi}{\partial \rho} dE.$$

Since *dE* is equal to the transferred heat, and the process is reversible we also have:

$$dS = \frac{1}{T} dE.$$

Equating formulas gives:

$$\frac{\partial \phi}{\partial \rho} = \frac{1}{T}.$$

This is the black body radiation law. So it's possible to determine the black body radiation from the function ϕ. Conversely, through integration one can obtain ϕ from the black body radiation law keeping in mind that ϕ vanishes for $\varrho = 0$.

Limiting law for the entropy of monochromatic radiation at low radiation density

Admittedly, the observations of "black body radiation" so far indicate that the law that Mr. Wien originally devised for the "black body radiation"

$$\rho = \alpha\nu^3 e^{-\beta\frac{\nu}{T}}$$

is not exactly valid. However, for large values of ν/T experiment completely confirms the law. We shall base our calculations on this formula, keeping in mind that the results will be valid within certain limitations only.

First, we get from this equation:

$$\frac{1}{T} = -\frac{1}{\beta\nu} \lg \frac{\rho}{\alpha\nu^3}$$

and then, using the relation obtained in the preceding section:

$$\phi(\rho, \nu) = -\frac{\rho}{\beta\nu} \left\{ \lg \frac{\rho}{\alpha\nu^3} - 1 \right\}.$$

Let there be a radiation of energy E, with a frequency between ν and $\nu + d\nu$. Let the radiation extend over volume v. The entropy of this radiation is:

$$S = v\phi(\rho, \nu) d\nu = -\frac{E}{\beta\nu} \left\{ \lg \frac{E}{v\alpha\nu^3 d\nu} - 1 \right\}.$$

We will limit ourselves to investigating the dependency of the radiation's entropy on the volume that is occupied. Let the entropy of the radiation be called S_0 when it occupies the volume v_0, then we get:

$$S - S_0 = \frac{E}{\beta\nu} \lg \left(\frac{v}{v_0} \right).$$

This equation shows that the entropy of monochromatic radiation of sufficiently low density varies with volume according to the same law as the entropy of an ideal gas or that of a dilute solution. In the following the equation just found will be interpreted in terms of the principle introduced by Mr. Boltzmann that says that the entropy of a system is a function of the probability of its state.

Molecular Theoretical investigation of the Volume Dependence of the Entropy of Gases and Dilute Solutions

In calculating Entropy on the grounds of molecular theory the word "probability" is often used in a meaning that isn't covered by the definition in probability theory. Especially the "cases of equal probability" are often set by hypothesis, where the applied theoretical representation is sufficiently definite to deduce probabilities without fixing them by hypothesis. I will show in a separate work that in considerations of thermal processes one obtains a complete result with the so-called "statistical probability". This way I hope to remove a logical difficulty that is in the way of fully implementing Boltzmann's principle. Here however only its general formulation and application in quite specific cases will be given.

When it's meaningful to talk about the probability of a state of a system, and additionally every increase of entropy can be described as a transition to a more probable state, the entropy S_1 of a system is a function of the probability W_1 of its instantaneous state. In the case of two systems S_1 and S_2, one can state:

$$S_1 = \phi_1(W_1),$$
$$S_2 = \phi_2(W_2).$$

If one considers these systems as a single system with entropy S and probability W, then:

$$S = S_1 + S_2 = \phi(W)$$

and

$$W = W_1 \cdot W_2.$$

The latter equation expresses that the states of the two systems are independent.

From these equations it follows:

$$\phi(W_1 \cdot W_2) = \phi_1(W_1) + \phi_2(W_2)$$

and hence finally

$$\begin{aligned} \phi_1(W_1) &= C \lg(W_1) + const.\ , \\ \phi_2(W_2) &= C \lg(W_2) + const.\ , \\ \phi(W) &= C \lg(W) + const. \end{aligned}$$

The quantity C is also a universal constant; it follows from kinetic gas theory, where the constants R and N have the same meaning as above. Denoting the entropy at a particular starting state as S_0, and the relative probability of a state with entropy S as W we have in general:

$$S - S_0 = \frac{R}{N} \lg W.$$

We now consider the following special case. Let a number (n) of movable points (for example molecules) be present in a volume v_0, these points will be the subject of our considerations. Other than these, arbitrarily many other movable points can be present. As to the law that describes how the considered points move around in the space the only assumption is that no part of the space (and no direction) is favored over others. The number of the (first-mentioned) points that we are considering is so small that mutual interactions are negligible.

The system considered, which can be for example an ideal gas or a diluted solution, has a certain entropy. We take a part of the volume v_0 with a size of v and we think of all n movable points displaced to that volume v, with otherwise no change of the system. Clearly this state has another entropy (S), and here we want to determine that entropy difference with the help of Boltzmann's principle.

We ask: how large is the probability of the last-mentioned state relative to the original state? Or, what is the probability that at some point in time all n independently moving points in a volume v_0 have by chance ended up in the volume v?

For this probability, which is a "statistical probability" one obtains the value:

$$W = \left(\frac{v}{v_0}\right)^n \quad ;$$

one derives from this, applying Boltzmann's principle:

$$S - S_0 = R\left(\frac{n}{N}\right) \lg\left(\frac{v}{v_0}\right).$$

It's noteworthy that for this derivation, from which the Boyle-Gay-Lussac law and the identical law of osmotic pressure can be easily derived thermodynamically [6], there is no need to make any assumption regarding the way the molucules move.

Interpretation of the Volume Dependence of the Entropy of Monochromatic Radiation using Boltzmann's Principle

In paragraph 4 we found for the dependence of Entropy of the monochromatic radiation on volume the expression:

$$S - S_0 = \frac{E}{\beta\nu} \lg\left(\frac{v}{v_0}\right).$$

This formula can be recast as follows:

$$S - S_0 = \frac{R}{N} \lg\left[\left(\frac{v}{v_0}\right)^{\frac{N}{R}\frac{E}{\beta\nu}}\right]$$

Comparing this with the general formula that expresses Boltzmann's principle

$$S - S_0 = \frac{R}{N} \lg W,$$

we arrive at the following conclusion:

If monochromatic radiation of frequency ν and energy E is enclosed (by reflecting walls) in the volume v_0, then the probability that at an arbitrary point in time all of the radiation energy located in a part v of the volume v_0 is:

$$W = \left(\frac{v}{v_0}\right)^{\frac{N}{R}\frac{E}{\beta\nu}} .$$

Subsequently we conclude:

In terms of heat theory monochromatic radiation of low density (within the realm of validity of Wien's radiation formula) behaves as if it consisted of independent energy quanta of the magnitude $R\beta\nu/N$.

We also want to compare the average magnitude of the energy quanta of the "black body radiation" with the mean average energy of the center-of-mass-motion of a molecule at the same temperature. The latter is $^3/_2(R/N)$T, and for the average energy of the Energy quanta Wien's formula gives:

$$\frac{\int\limits_0^\infty \alpha\nu^3 e^{-\frac{\beta\nu}{T}} d\nu}{\int\limits_0^\infty \frac{N}{R\beta\nu}\alpha\nu^3 e^{-\frac{\beta\nu}{T}} d\nu} = 3\frac{R}{N}T.$$

The fact that monochromatic radiation (of sufficiently low density) behaves as regards to dependency of entropy on volume like a discontinuous medium that consists of energy quanta of magnitude $R\beta\nu/N$ suggests we should investigate whether the laws of generation and transformation of light are what they must be if light consisted of such energy quanta. In the following we will address that question.

Stokes' Rule

Let monochromatic light be transformed by photoluminence into light of another frequency, and let it be assumed that according to the result just obtained the generating as well as the generated light consists of energy quanta of magnitude $(R/N)\beta\nu$, where ν is the corresponding frequency. The transformation process can then be interpreted as follows. Each generating energy quantum of frequency ν_1 is absorbed and generates—at least with sufficiently small density of the generating energy quanta—by itself a light quantum of of frequency ν_2; possibly other light quanta of frequency ν_3, ν_4 etc. as well as other form of energy (e.g heat) can be generated simultaneously. Through which intermedia processes the final result comes about is immaterial. If the photoluminescing substance isn't a continuous source of energy it follows from the energy principle that the energy of the generated energy quanta are not larger than the generating light quanta; therefore the following relation must hold:

$$\frac{R}{N}\beta\nu_2 \leqq \frac{R}{N}\beta\nu_1$$

or

$$\nu_2 \leqq \nu_1 .$$

As is well known this is Stokes' rule.

Especially noteworthy is that with weak illumination the amount of generated light must, other circumstances being equal, be proportional to the amount of exciting light, since every incident energy quantum will cause one elementary process of the above indicated kind, independent of the action of other exciting energy quanta. In particular there will be no lower limit of the intensity of the exciting light below which the light would be incapable of exciting light.

According to the way the understanding of the phenomena is laid down here deviations from Stokes' rule are conceivable in the following cases:

1. When the number of energy quanta per unit of volume that are simultaneously involved in the transformation is so large that the energy quantum of the generated light can receive the energy of several exciting energy quanta.
2. When the generating (or generated) light does not have the energy characteristics of "black body radiation" that is in the realm of validity of Wien's law, when for instance the exciting light is generated by a body of such high temperature that for the wavelengths considered Wien's law is no longer valid.

The last mentioned possibility merits special attention. According to the developed understanding it cannot be excluded that a "non-Wienian radiation", even in high dilution, would behave energetically differently from a "black body radiation" within the validity range of Wien's law.

On the Generation of Cathode Rays by Illumination of Solid Bodies

The usual understanding, that the energy of light is distributed over the space through which it travels in a continuous way encounters extraordinarily large difficulties in attempts to explain photo-electric phenomena, as has been presented in the groundbreaking article by Mr. Lenard. [7].

According to the understanding that the exciting light consists of energy quanta of energy $(R/N)\beta\nu$ the generation of cathode rays by light can be conceived as follows. Quanta of energy penetrate the surface layer of the solid, and their energy is transformed, at least partially, in kinetic energy of electrons. The simplest picture is one where the light quantum gives its entire energy to a single electron; we assume that this will occur. However, it must not be excluded that electrons accept the energy of light quanta only partially. An electron that has been loaded with kinetic energy will have lost some of its energy when it arrives at the surface. Other than that we must assume that on leaving the solid every electron must do an amount of work P (characteristic of that solid). Electrons residing right at the surface, excited at right angles to it, will leave the solid with the largest normal velocity. The kinetic energy of such electrons is

$$\frac{R}{N}\beta\nu - P.$$

If the body is charged to a positive potential Π and surrounded by conductors with potential zero and Π is just enough to prevent loss of electricity by the body, then we must have:

$$\Pi\epsilon = \frac{R}{N}\beta\nu - P,$$

where ε is the electrical mass of the electron, or

$$\Pi E = R\beta\nu - P',$$

where E is the charge of one gram equivalent of a single-valued ion and P' is the potentel of this amount of negative electricity with respect to this body. [8]

If we set $E = 9.6 \cdot 10^3$, then $\Pi \cdot 10^{-8}$ is the potential in volts that the body will attain when it is irradiated in vacuum.

To see now whether the derived relation agrees with experiment to within an order of magnitude we set $P' = 0$, $\nu = 1.03 \cdot 10^{15}$ (corresponding to the ultraviolet limit of the solar spectrum), and $\beta = 4.866 \cdot 10^{-11}$. We obtain $\Pi \cdot 10^7 = 4.3$ Volt, which agrees to within an order of magnitude with the results of Mr. Lenard. [9]

If the formula derived is correct, then Π, as a function of frequency of the excited light represented in Cartesian coordinates, must be a straight line, whose inclination is independent from the nature of the substance investigated.

As far as I can see no contradiction exists between our understanding and the properties of photo-electric action observed by Mr. Lenard. If each energy quantum of the exciting light releases its energy independently from all others to the electrons, the distribution of velocities of the electrons, which means the quality of the generated cathode radiation, will be independent of the intensity of the exciting light; the number of electrons that exits the body, on the other hand, will, in otherwise equal circumstances, be proportional to the intensity of the exciting light. [10]

We expect that limits of validity of these rules will be similar in nature to the expected deviations from Stokes' rule.

In the preceding it has been assumed that the energy of at least some of the energy quanta of the generating light is transferred completely to a single electron. If one does not start with that natural supposition then instead of the above equation one obtains:

$$\Pi E + P' \leqq R\beta\nu.$$

For cathode-luminescence, which constitutes the inverse process of the one just examined, on obtains by way of analogous consideration:

$$\Pi E + P' \geqq R\beta\nu.$$

For the materials investigated by Mr. Lenard PE is always significantly larger than $R\beta\nu$, as the voltage that the cathode rays have had to traverse to generate even visible light is in some cases several hundred, in other cases thousands of volts. [11]

Ionization of Gases by Ultraviolet Light

We have to assume that in ionization of a gas by ultraviolet light always one absorbed light energy quantum is used for the ionization of just one gas molecule. Firstly it follows that the ionization energy (that is, the theoretically necessary energy to ionize) of a molecule cannot be larger than the energy of an absorbed light energy quantum. Taking J as the (theoretical) ionization energy per gram equivalent, we have:

$$R\beta\nu \geqq J.$$

According to Lenard's measurements for air the largest wavelength that has an effect is about $1.9 \cdot 10^{-5}$ cm, so

$$R\beta\nu = 6.4 \cdot 10^{12} \text{ Erg} \geqq J.$$

An upper limit for the ionization energy can also be obtained from the ionization voltage in rarefied gases. According to Stark [12] the smallest measured ionization voltage (for platinum anodes) is for air about 10 volt. [13] We have thus for J an upper limit $9.6 \cdot 10^{12}$, which is nearly the same as the one just found. There is another consequence that in my mind is very important to verify. If every light energy quantum ionizes one molecule then the following relation must exist between the absorbed quantity of light L and the number j of thereby ionized gram molecules:

$$j = \frac{L}{R\beta\nu}.$$

If our understanding reflects reality this relation must hold for every gas that (at the particular frequency) has no absorption that isn't accompanied by ionization.

Bern, march 17, 1905

---

1. ↑ This assumption is equivalent to the condition that the mean kinetic energies of gas molecules and electrons are equal to each other when there is thermal equilibrium. As is known, using this condition Mr. Drude has theoretically derived the relation between thermal and electric conductivity of metals.
2. ↑ M. Planck, Ann. d. Phys. **1** p.99. 1900.
3. ↑ This condition can be formulated as follows. We expand the Z-component of the electric force (Z) in a given point in the space between the time coordinates of t=0 and t=T (where T is a large amount of time compared to all the vibration periods considered) in a Fourier series

   $$Z = \sum_{\nu=1}^{\nu=\infty} A_\nu sin\left(2\pi\nu\frac{t}{T} + \alpha_\nu\right) \;,$$

   where $A_\nu \geqq 0$ and $0 \leqq a_\nu \leqq 2\pi$. Performing this expansion arbitrarily often with arbitrarily chosen initial times yields a range of different combinations for the quantities A_ν and α_ν. Then for the frequencies of the different combinations of the quantities A_ν and α_ν there are the (statistical) probabilities dW of the form:

   $$dW = f(A_1\ A_2, \ldots \alpha_1\ \alpha_2 \ldots) dA_1 dA_2, \ldots d\alpha_1\ d\alpha_2 \ldots$$

The radiation is then as unordered as imaginable, if

$$f(A_1, A_2 \ \ldots \alpha_1, \alpha_2 \ldots) = F_1(A_1)F_2(A_2)\ldots f_1(\alpha_1).f_2(\alpha_2)\ldots$$

That is if the probability of a particular value of A and α respectively is independent of the value of other values of A and x respectively. The more closely the demand is satisfied that the separate pairs of values A_ν and α_ν depend on the emission and absorption process of *separate* resonators, the more closely will the examined case be one of being as unordered as imaginable.

4. ↑ M. Planck, Ann. d. Phys. 4. p.561. 1901.
5. ↑ This is an arbitrary assumption. The natural course of action is to stay with this simplest assumption until experiment forces us to abandon it.
6. ↑ If E is the energy of the system, then one obtains:

$$-d(E-TS) = pdv = TdS = TR\frac{n}{N}\frac{dv}{v};$$

therefore

$$pv = R\frac{n}{N}T.$$

7. ↑ P. Lenard, Ann. d. Phys. 8. p.169 u. 170. 1902.
8. ↑ If one assumes that in order to release an electron from a neutral molecule light must do a certain amount of work then one doesn't have to change the derived relation; one only has to think of P' as the sum of two terms.
9. ↑ P. Lenard, Ann. d. Phys. 8. p165. u. 184 Taf. I, Fig.2 1902.
10. ↑ P. Lenard, l. c. p.150 und p. 166-168.
11. ↑ P. Lenard, Ann. d. Phys. **12.** p.469. 1903.
12. ↑ J. Stark, Die Elektricität in Gasen p. 57. Leipzig 1902.
13. ↑ within the gas the ionization voltage for negative ions is nonetheless five times larger

REPRINTED FROM:

THE COLLECTED PAPERS OF

Albert Einstein

VOLUME 6

THE BERLIN YEARS:
WRITINGS, 1914–1917

A. J. Kox, Martin J. Klein, and Robert Schulmann
EDITORS

József Illy and Jean Eisenstaedt
CONTRIBUTING EDITORS

Rita Fountain and Annette Pringle
EDITORIAL ASSISTANTS

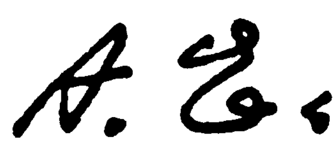
ENGLISH TRANSLATION OF SELECTED TEXTS

Alfred Engel, TRANSLATOR
Engelbert Schucking, CONSULTANT

DOC. 30

The Foundation of the General Theory of Relativity

(pp. 146–200 in translation volume)



Princeton University Press
1997

Doc. 30

[p. 769] **The Foundation of the General Theory of Relativity**

by A. Einstein

[This first page was missing in the existing translation.]

The theory which is presented in the following pages conceivably constitutes the farthest-reaching generalization of a theory which, today, is generally called the
[1] "theory of relativity"; I will call the latter one—in order to distinguish it from the
[2] first named—the "special theory of relativity," which I assume to be known. The generalization of the theory of relativity has been facilitated considerably by Minkowski, a mathematician who was the first one to recognize the formal
[3] equivalence of space coordinates and the time coordinate, and utilized this in the construction of the theory. The mathematical tools that are necessary for general relativity were readily available in the "absolute differential calculus," which is based upon the research on non-Euclidean manifolds by Gauss, Riemann, and Christoffel, and which has been systematized by Ricci and Levi-Civita and has already been
[4] applied to problems of theoretical physics. In section B of the present paper I developed all the necessary mathematical tools—which cannot be assumed to be known to every physicist—and I tried to do it in as simple and transparent a manner as possible, so that a special study of the mathematical literature is not required for
[5] the understanding of the present paper. Finally, I want to acknowledge gratefully my friend, the mathematician Grossmann, whose help not only saved me the effort of
[6] studying the pertinent mathematical literature, but who also helped me in my search for the field equations of gravitation.

[The balance of this translation is reprinted from H. A. Lorentz et al., *The Principle of Relativity*, trans. W. Perrett and G. B. Jeffery (Methuen, 1923; Dover rpt., 1952).]

THE FOUNDATION OF THE GENERAL THEORY OF RELATIVITY

By A. EINSTEIN

A. Fundamental Considerations on the Postulate of Relativity

§ 1. Observations on the Special Theory of Relativity

THE special theory of relativity is based on the following postulate, which is also satisfied by the mechanics of Galileo and Newton.

If a system of co-ordinates K is chosen so that, in relation to it, physical laws hold good in their simplest form, the *same* laws also hold good in relation to any other system of co-ordinates K′ moving in uniform translation relatively to K. This postulate we call the "special principle of relativity." The word "special" is meant to intimate that the principle is restricted to the case when K′ has a motion of uniform translation relatively to K, but that the equivalence of K′ and K does not extend to the case of non-uniform motion of K′ relatively to K.

Thus the special theory of relativity does not depart from classical mechanics through the postulate of relativity, but through the postulate of the constancy of the velocity of light *in vacuo*, from which, in combination with the special principle of relativity, there follow, in the well-known way, the relativity of simultaneity, the Lorentzian transformation, and the related laws for the behaviour of moving bodies and clocks.

The modification to which the special theory of relativity has subjected the theory of space and time is indeed far-reaching, but one important point has remained unaffected.

For the laws of geometry, even according to the special theory of relativity, are to be interpreted directly as laws relating to the possible relative positions of solid bodies at rest; and, in a more general way, the laws of kinematics are to be interpreted as laws which describe the relations of measuring bodies and clocks. To two selected material points of a stationary rigid body there always corresponds a distance of quite definite length, which is independent of the locality and orientation of the body, and is also independent of the time. To two selected positions of the hands of a clock at rest relatively to the privileged system of reference there always corresponds an interval of time of a definite length, which is independent of place and time. We shall soon see that the general theory of relativity cannot adhere to this simple physical interpretation of space and time.

§ 2. The Need for an Extension of the Postulate of Relativity

In classical mechanics, and no less in the special theory of relativity, there is an inherent epistemological defect which
[7] was, perhaps for the first time, clearly pointed out by Ernst Mach. We will elucidate it by the following example:—Two fluid bodies of the same size and nature hover freely in space at so great a distance from each other and from all other masses that only those gravitational forces need be taken into account which arise from the interaction of different parts of the same body. Let the distance between the two bodies be invariable, and in neither of the bodies let there be any relative movements of the parts with respect to one another. But let either mass, as judged by an observer at rest relatively to the other mass, rotate with constant angular velocity about the line joining the masses. This is a verifiable relative motion of the two bodies. Now let us imagine that each of the bodies has been surveyed by means of measuring instruments at rest relatively to itself, and let the surface of S_1 prove to be a sphere, and that of S_2 an ellipsoid of revolution. Thereupon we put the question—What is the reason for this difference in the two bodies? No answer can

be admitted as epistemologically satisfactory,* unless the reason given is an *observable fact of experience.* The law of causality has not the significance of a statement as to the world of experience, except when *observable facts* ultimately appear as causes and effects.

Newtonian mechanics does not give a satisfactory answer to this question. It pronounces as follows:—The laws of mechanics apply to the space R_1, in respect to which the body S_1 is at rest, but not to the space R_2, in respect to which the body S_2 is at rest. But the privileged space R_1 of Galileo, thus introduced, is a merely *factitious* cause, and not a thing that can be observed. It is therefore clear that Newton's mechanics does not really satisfy the requirement of causality in the case under consideration, but only apparently does so, since it makes the factitious cause R_1 responsible for the observable difference in the bodies S_1 and S_2.

The only satisfactory answer must be that the physical system consisting of S_1 and S_2 reveals within itself no imaginable cause to which the differing behaviour of S_1 and S_2 can be referred. The cause must therefore lie *outside* this system. We have to take it that the general laws of motion, which in particular determine the shapes of S_1 and S_2, must be such that the mechanical behaviour of S_1 and S_2 is partly conditioned, in quite essential respects, by distant masses which we have not included in the system under consideration. These distant masses and their motions relative to S_1 and S_2 must then be regarded as the seat of the causes (which must be susceptible to observation) of the different behaviour of our two bodies S_1 and S_2. They take over the rôle of the factitious cause R_1. Of all imaginable spaces R_1, R_2, etc., in any kind of motion relatively to one another, there is none which we may look upon as privileged *a priori* without reviving the above-mentioned epistemological objection. *The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion.* Along this road we arrive at an extension of the postulate of relativity.

In addition to this weighty argument from the theory of

* Of course an answer may be satisfactory from the point of view of epistemology, and yet be unsound physically, if it is in conflict with other experiences.

knowledge, there is a well-known physical fact which favours an extension of the theory of relativity. Let K be a Galilean system of reference, i.e. a system relatively to which (at least in the four-dimensional region under consideration) a mass, sufficiently distant from other masses, is moving with uniform motion in a straight line. Let K′ be a second system of reference which is moving relatively to K in *uniformly accelerated* translation. Then, relatively to K′, a mass sufficiently distant from other masses would have an accelerated motion such that its acceleration and direction of acceleration are independent of the material composition and physical state of the mass.

Does this permit an observer at rest relatively to K′ to infer that he is on a "really" accelerated system of reference? The answer is in the negative; for the above-mentioned relation of freely movable masses to K′ may be interpreted equally well in the following way. The system of reference K′ is unaccelerated, but the space-time territory in question is under the sway of a gravitational field, which generates the
[8] accelerated motion of the bodies relatively to K′.

This view is made possible for us by the teaching of experience as to the existence of a field of force, namely, the gravitational field, which possesses the remarkable property of imparting the same acceleration to all bodies.* The mechanical behaviour of bodies relatively to K′ is the same as presents itself to experience in the case of systems which we are wont to regard as "stationary" or as "privileged." Therefore, from the physical standpoint, the assumption readily suggests itself that the systems K and K′ may both with equal right be looked upon as "stationary," that is to say, they have an equal title as systems of reference for the physical description of phenomena.

It will be seen from these reflexions that in pursuing the general theory of relativity we shall be led to a theory of gravitation, since we are able to "produce" a gravitational field merely by changing the system of co-ordinates. It will also be obvious that the principle of the constancy of the velocity of light *in vacuo* must be modified, since we easily

[9] * Eötvös has proved experimentally that the gravitational field has this property in great accuracy.

recognize that the path of a ray of light with respect to K′ must in general be curvilinear, if with respect to K light is propagated in a straight line with a definite constant velocity.

§ 3. The Space-Time Continuum. Requirement of General Co-Variance for the Equations Expressing General Laws of Nature

In classical mechanics, as well as in the special theory of relativity, the co-ordinates of space and time have a direct physical meaning. To say that a point-event has the X_1 co-ordinate x_1 means that the projection of the point-event on the axis of X_1, determined by rigid rods and in accordance with the rules of Euclidean geometry, is obtained by measuring off a given rod (the unit of length) x_1 times from the origin of co-ordinates along the axis of X_1. To say that a point-event has the X_4 co-ordinate $x_4 = t$, means that a standard clock, made to measure time in a definite unit period, and which is stationary relatively to the system of co-ordinates and practically coincident in space with the point-event,* will have measured off $x_4 = t$ periods at the occurrence of the event.

This view of space and time has always been in the minds of physicists, even if, as a rule, they have been unconscious of it. This is clear from the part which these concepts play in physical measurements; it must also have underlain the reader's reflexions on the preceding paragraph (§ 2) for him to connect any meaning with what he there read. But we shall now show that we must put it aside and replace it by a more general view, in order to be able to carry through the postulate of general relativity, if the special theory of relativity applies to the special case of the absence of a gravitational field.

In a space which is free of gravitational fields we introduce a Galilean system of reference K (x, y, z, t), and also a system of co-ordinates K′ (x', y', z', t') in uniform rotation relatively to K. Let the origins of both systems, as well as their axes

* We assume the possibility of verifying "simultaneity" for events immediately proximate in space, or—to speak more precisely—for immediate proximity or coincidence in space-time, without giving a definition of this fundamental concept.

of Z, permanently coincide. We shall show that for a space-time measurement in the system K′ the above definition of the physical meaning of lengths and times cannot be maintained. For reasons of symmetry it is clear that a circle around the origin in the X, Y plane of K may at the same time be regarded as a circle in the X′, Y′ plane of K′. We suppose that the circumference and diameter of this circle have been measured with a unit measure infinitely small compared with the radius, and that we have the quotient of the two results. If this experiment were performed with a measuring-rod at rest relatively to the Galilean system K, the quotient would be π. With a measuring-rod at rest relatively to K′, the quotient would be greater than π. This is readily understood if we envisage the whole process of measuring from the "stationary" system K, and take into consideration that the measuring-rod applied to the periphery undergoes a Lorentzian contraction, while the one applied along the radius does not. Hence Euclidean geometry does not apply to K′. The notion of co-ordinates defined above, which presupposes the validity of Euclidean geometry, therefore breaks
[10] down in relation to the system K′. So, too, we are unable to introduce a time corresponding to physical requirements in K′, indicated by clocks at rest relatively to K′. To convince ourselves of this impossibility, let us imagine two clocks of identical constitution placed, one at the origin of co-ordinates, and the other at the circumference of the circle, and both envisaged from the "stationary" system K. By a familiar result of the special theory of relativity, the clock at the circumference—judged from K—goes more slowly than the other, because the former is in motion and the latter at rest. An observer at the common origin of co-ordinates, capable of observing the clock at the circumference by means of light, would therefore see it lagging behind the clock beside him. As he will not make up his mind to let the velocity of light along the path in question depend explicitly on the time, he will interpret his observations as showing that the clock at the circumference "really" goes more slowly than the clock at the origin. So he will be obliged to define time in such a way that the rate of a clock depends upon where the clock may be.

We therefore reach this result:—In the general theory of relativity, space and time cannot be defined in such a way that differences of the spatial co-ordinates can be directly measured by the unit measuring-rod, or differences in the time co-ordinate by a standard clock.

The method hitherto employed for laying co-ordinates into the space-time continuum in a definite manner thus breaks down, and there seems to be no other way which would allow us to adapt systems of co-ordinates to the four-dimensional universe so that we might expect from their application a particularly simple formulation of the laws of nature. So there is nothing for it but to regard all imaginable systems of co-ordinates, on principle, as equally suitable for the description of nature. This comes to requiring that:—

The general laws of nature are to be expressed by equations which hold good for all systems of co-ordinates, that is, are co-variant with respect to any substitutions whatever (generally co-variant).

It is clear that a physical theory which satisfies this postulate will also be suitable for the general postulate of relativity. For the sum of *all* substitutions in any case includes those which correspond to all relative motions of three-dimensional systems of co-ordinates. That this requirement of general co-variance, which takes away from space and time the last remnant of physical objectivity, is a natural one, will be seen from the following reflexion. All our space-time verifications invariably amount to a determination
of space-time coincidences. If, for example, events consisted [11]
merely in the motion of material points, then ultimately nothing would be observable but the meetings of two or more of these points. Moreover, the results of our measurings are nothing but verifications of such meetings of the material points of our measuring instruments with other material points, coincidences between the hands of a clock and points on the clock dial, and observed point-events happening at the same place at the same time.

The introduction of a system of reference serves no other purpose than to facilitate the description of the totality of such coincidences. We allot to the universe four space-time variables x_1, x_2, x_3, x_4 in such a way that for every point-event there is a corresponding system of values of the variables $x_1 \ldots x_4$. To two coincident point-events there corresponds one system of values of the variables $x_1 \ldots x_4$, i.e. coincidence is characterized by the identity of the co-ordinates. If, in place of the variables $x_1 \ldots x_4$, we introduce functions of them, x'_1, x'_2, x'_3, x'_4, as a new system of co-ordinates, so that the systems of values are made to correspond to one another without ambiguity, the equality of all four co-ordinates in the new system will also serve as an expression for the space-time coincidence of the two point-events. As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of co-ordinates to others, that is to say, we arrive at the requirement of general co-variance.

§ 4. The Relation of the Four Co-ordinates to Measurement in Space and Time

It is not my purpose in this discussion to represent the general theory of relativity as a system that is as simple and logical as possible, and with the minimum number of axioms; but my main object is to develop this theory in such a way that the reader will feel that the path we have entered upon is psychologically the natural one, and that the underlying assumptions will seem to have the highest possible degree of security. With this aim in view let it now be granted that:—

For infinitely small four-dimensional regions the theory of relativity in the restricted sense is appropriate, if the co-ordinates are suitably chosen.

For this purpose we must choose the acceleration of the infinitely small ("local") system of co-ordinates so that no gravitational field occurs; this is possible for an infinitely small region. Let X_1, X_2, X_3, be the co-ordinates of space, and X_4 the appertaining co-ordinate of time measured in the appropriate unit.* If a rigid rod is imagined to be given as the unit measure, the co-ordinates, with a given orientation of the system of co-ordinates, have a direct physical meaning

* The unit of time is to be chosen so that the velocity of light *in vacuo* as measured in the "local" system of co-ordinates is to be equal to unity.

in the sense of the special theory of relativity. By the special theory of relativity the expression

$$ds^2 = -dX_1^2 - dX_2^2 - dX_3^2 + dX_4^2 \quad . \quad . \quad (1)$$

then has a value which is independent of the orientation of the local system of co-ordinates, and is ascertainable by measurements of space and time. The magnitude of the linear element pertaining to points of the four-dimensional continuum in infinite proximity, we call ds. If the ds belonging to the element $dX_1 \ldots dX_4$ is positive, we follow Minkowski in calling it time-like; if it is negative, we call it space-like.

To the "linear element" in question, or to the two infinitely proximate point-events, there will also correspond definite differentials $dx_1 \ldots dx_4$ of the four-dimensional co-ordinates of any chosen system of reference. If this system, as well as the "local" system, is given for the region under consideration, the dX_ν will allow themselves to be represented here by definite linear homogeneous expressions of the dx_σ:—

$$dX_\nu = \sum_\sigma a_{\nu\sigma} dx_\sigma \quad . \quad . \quad . \quad . \quad (2)$$

Inserting these expressions in (1), we obtain

$$ds^2 = \sum_{\tau\sigma} g_{\sigma\tau} dx_\sigma dx_\tau, \quad . \quad . \quad . \quad . \quad (3)$$

where the $g_{\sigma\tau}$ will be functions of the x_σ. These can no longer be dependent on the orientation and the state of motion of the "local" system of co-ordinates, for ds^2 is a quantity ascertainable by rod-clock measurement of point-events infinitely proximate in space-time, and defined independently of any particular choice of co-ordinates. The $g_{\sigma\tau}$ are to be chosen here so that $g_{\sigma\tau} = g_{\tau\sigma}$; the summation is to extend over all values of σ and τ, so that the sum consists of 4×4 terms, of which twelve are equal in pairs.

The case of the ordinary theory of relativity arises out of the case here considered, if it is possible, by reason of the particular relations of the $g_{\sigma\tau}$ in a finite region, to choose the system of reference in the finite region in such a way that the $g_{\sigma\tau}$ assume the constant values

$$\left.\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{matrix}\right\} \quad . \quad . \quad . \quad (4)$$

We shall find hereafter that the choice of such co-ordinates is, in general, not possible for a finite region.

From the considerations of § 2 and § 3 it follows that the quantities $g_{\tau\sigma}$ are to be regarded from the physical standpoint as the quantities which describe the gravitational field in relation to the chosen system of reference. For, if we now assume the special theory of relativity to apply to a certain four-dimensional region with the co-ordinates properly chosen, then the $g_{\sigma\tau}$ have the values given in (4). A free material point then moves, relatively to this system, with uniform motion in a straight line. Then if we introduce new space-time co-ordinates x_1, x_2, x_3, x_4, by means of any substitution we choose, the $g^{\sigma\tau}$ in this new system will no longer be constants, but functions of space and time. At the same time the motion of the free material point will present itself in the new co-ordinates as a curvilinear non-uniform motion, and the law of this motion will be independent of the nature of the moving particle. We shall therefore interpret this motion as a motion under the influence of a gravitational field. We thus find the occurrence of a gravitational field connected with a space-time variability of the g_{σ}. So, too, in the general case, when we are no longer able by a suitable choice of co-ordinates to apply the special theory of relativity to a finite region, we shall hold fast to the view that the $g_{\sigma\tau}$ describe the gravitational field.

Thus, according to the general theory of relativity, gravitation occupies an exceptional position with regard to other forces, particularly the electromagnetic forces, since the ten functions representing the gravitational field at the same time define the metrical properties of the space measured.

B. Mathematical Aids to the Formulation of Generally Covariant Equations

Having seen in the foregoing that the general postulate of relativity leads to the requirement that the equations of physics shall be covariant in the face of any substitution of the co-ordinates $x_1 \ldots x_4$, we have to consider how such generally covariant equations can be found. We now turn to this purely mathematical task, and we shall find that in its solution a fundamental rôle is played by the invariant ds given in equation (3), which, borrowing from Gauss's theory of surfaces, we have called the "linear element."

The fundamental idea of this general theory of covariants is the following:—Let certain things ("tensors") be defined with respect to any system of co-ordinates by a number of functions of the co-ordinates, called the "components" of the tensor. There are then certain rules by which these components can be calculated for a new system of co-ordinates, if they are known for the original system of co-ordinates, and if the transformation connecting the two systems is known. The things hereafter called tensors are further characterized by the fact that the equations of transformation for their components are linear and homogeneous. Accordingly, all the components in the new system vanish, if they all vanish in the original system. If, therefore, a law of nature is expressed by equating all the components of a tensor to zero, it is generally covariant. By examining the laws of the formation of tensors, we acquire the means of formulating generally covariant laws.

§ 5. Contravariant and Covariant Four-vectors

Contravariant Four-vectors.—The linear element is defined by the four "components" dx_ν, for which the law of transformation is expressed by the equation

$$dx'_\sigma = \sum_\nu \frac{\partial x'_\sigma}{\partial x_\nu} dx_\nu \quad . \quad . \quad . \quad . \quad (5)$$

The dx'_σ are expressed as linear and homogeneous functions of the dx_ν. Hence we may look upon these co-ordinate differentials as the components of a "tensor" of the particular kind which we call a contravariant four-vector. Any thing which is defined relatively to the system of co-ordinates by four quantities A^ν, and which is transformed by the same law

$$A'^\sigma = \sum_\nu \frac{\partial x'_\sigma}{\partial x_\nu} A^\nu, \quad . \quad . \quad . \quad . \quad (5a)$$

we also call a contravariant four-vector. From (5a) it follows at once that the sums $A^\sigma \pm B^\sigma$ are also components of a four-vector, if A^σ and B^σ are such. Corresponding relations hold for all "tensors" subsequently to be introduced. (Rule for the addition and subtraction of tensors.)

Covariant Four-vectors.—We call four quantities A_ν the components of a covariant four-vector, if for any arbitrary choice of the contravariant four-vector B^ν

$$\sum_\nu A_\nu B^\nu = \text{Invariant} \quad . \quad . \quad . \quad (6)$$

The law of transformation of a covariant four-vector follows from this definition. For if we replace B^ν on the right-hand side of the equation

$$\sum_\sigma A'_\sigma B'^\sigma = \sum_\nu A_\nu B^\nu$$

by the expression resulting from the inversion of (5a),

$$\sum_\sigma \frac{\partial x_\nu}{\partial x'_\sigma} B'^\sigma,$$

we obtain

$$\sum_\sigma B'^\sigma \sum_\nu \frac{\partial x_\nu}{\partial x'_\sigma} A_\nu = \sum_\sigma B'^\sigma A'_\sigma.$$

Since this equation is true for arbitrary values of the B'^σ, it follows that the law of transformation is

$$A'_\sigma = \sum_\nu \frac{\partial x_\nu}{\partial x'_\sigma} A_\nu \quad . \quad . \quad . \quad . \quad (7)$$

Note on a Simplified Way of Writing the Expressions.—A glance at the equations of this paragraph shows that there is always a summation with respect to the indices which occur twice under a sign of summation (e.g. the index ν in (5)), and only with respect to indices which occur twice. It is therefore possible, without loss of clearness, to omit the sign of summation. In its place we introduce the convention:—If an index occurs twice in one term of an expression, it is
[12] always to be summed unless the contrary is expressly stated.

The difference between covariant and contravariant four-vectors lies in the law of transformation ((7) or (5) respectively). Both forms are tensors in the sense of the general remark above. Therein lies their importance. Following Ricci and

Levi-Civita, we denote the contravariant character by placing the index above, the covariant by placing it below. [13]

§ 6. Tensors of the Second and Higher Ranks

Contravariant Tensors.—If we form all the sixteen products $A^{\mu\nu}$ of the components A^μ and B^ν of two contravariant four-vectors

$$A^{\mu\nu} = A^\mu B^\nu \quad . \quad . \quad . \quad . \quad (8)$$

then by (8) and (5a) $A^{\mu\nu}$ satisfies the law of transformation

$$A'^{\sigma\tau} = \frac{\partial x'_\sigma}{\partial x_\mu}\frac{\partial x'_\tau}{\partial x_\nu} A^{\mu\nu} \quad . \quad . \quad . \quad (9)$$

We call a thing which is described relatively to any system of reference by sixteen quantities, satisfying the law of transformation (9), a contravariant tensor of the second rank. Not every such tensor allows itself to be formed in accordance with (8) from two four-vectors, but it is easily shown that any given sixteen $A^{\mu\nu}$ can be represented as the sums of the $A^\mu B^\nu$ of four appropriately selected pairs of four-vectors. Hence we can prove nearly all the laws which apply to the tensor of the second rank defined by (9) in the simplest manner by demonstrating them for the special tensors of the type (8).

Contravariant Tensors of Any Rank.—It is clear that, on the lines of (8) and (9), contravariant tensors of the third and higher ranks may also be defined with 4^3 components, and so on. In the same way it follows from (8) and (9) that the contravariant four-vector may be taken in this sense as a contravariant tensor of the first rank.

Covariant Tensors.—On the other hand, if we take the sixteen products $A_{\mu\nu}$ of two covariant four-vectors A_μ and B_ν,

$$A_{\mu\nu} = A_\mu B_\nu, \quad . \quad . \quad . \quad . \quad (10)$$

the law of transformation for these is

$$A'_{\sigma\tau} = \frac{\partial x_\mu}{\partial x'_\sigma}\frac{\partial x_\nu}{\partial x'_\tau} A_{\mu\nu} \quad . \quad . \quad . \quad (11)$$

This law of transformation defines the covariant tensor of the second rank. All our previous remarks on contravariant tensors apply equally to covariant tensors.

NOTE.—It is convenient to treat the scalar (or invariant) both as a contravariant and a covariant tensor of zero rank.

Mixed Tensors.—We may also define a tensor of the second rank of the type

$$A_\mu^\nu = A_\mu B^\nu \quad . \quad . \quad . \quad . \quad (12)$$

which is covariant with respect to the index μ, and contravariant with respect to the index ν. Its law of transformation is

$$A'^{\tau}_{\sigma} = \frac{\partial x'_\tau}{\partial x_\nu}\frac{\partial x_\mu}{\partial x'_\sigma}A^\nu_\mu \quad . \quad . \quad . \quad (13)$$

Naturally there are mixed tensors with any number of indices of covariant character, and any number of indices of contravariant character. Covariant and contravariant tensors may be looked upon as special cases of mixed tensors.

Symmetrical Tensors.—A contravariant, or a covariant tensor, of the second or higher rank is said to be symmetrical if two components, which are obtained the one from the other by the interchange of two indices, are equal. The tensor $A^{\mu\nu}$, or the tensor $A_{\mu\nu}$, is thus symmetrical if for any combination of the indices μ, ν,

$$A^{\mu\nu} = A^{\nu\mu}, \quad . \quad . \quad . \quad . \quad (14)$$

or respectively,

$$A_{\mu\nu} = A_{\nu\mu}. \quad . \quad . \quad . \quad . \quad (14a)$$

It has to be proved that the symmetry thus defined is a property which is independent of the system of reference. It follows in fact from (9), when (14) is taken into consideration, that

$$A'^{\sigma\tau} = \frac{\partial x'_\sigma}{\partial x_\mu}\frac{\partial x'_\tau}{\partial x_\nu}A^{\mu\nu} = \frac{\partial x'_\sigma}{\partial x_\mu}\frac{\partial x'_\tau}{\partial x_\nu}A^{\nu\mu} = \frac{\partial x'_\sigma}{\partial x_\nu}\frac{\partial x'_\tau}{\partial x_\mu}A^{\mu\nu} = A'^{\tau\sigma}.$$

The last equation but one depends upon the interchange of the summation indices μ and ν, i.e. merely on a change of notation.

Antisymmetrical Tensors.—A contravariant or a covariant tensor of the second, third, or fourth rank is said to be antisymmetrical if two components, which are obtained the one from the other by the interchange of two indices, are equal and of opposite sign. The tensor $A^{\mu\nu}$, or the tensor $A_{\mu\nu}$, is therefore antisymmetrical, if always

$$A^{\mu\nu} = -A^{\nu\mu}, \quad . \quad . \quad . \quad . \quad (15)$$

or respectively,

$$A_{\mu\nu} = -A_{\nu\mu} \quad . \quad . \quad . \quad . \quad (15a)$$

Of the sixteen components $A^{\mu\nu}$, the four components $A^{\mu\mu}$ vanish; the rest are equal and of opposite sign in pairs, so that there are only six components numerically different (a six-vector). Similarly we see that the antisymmetrical tensor of the third rank $A^{\mu\nu\sigma}$ has only four numerically different components, while the antisymmetrical tensor $A^{\mu\nu\sigma\tau}$ has only one. There are no antisymmetrical tensors of higher rank than the fourth in a continuum of four dimensions.

§ 7. Multiplication of Tensors

Outer Multiplication of Tensors.—We obtain from the components of a tensor of rank n and of a tensor of rank m the components of a tensor of rank $n + m$ by multiplying each component of the one tensor by each component of the other. Thus, for example, the tensors T arise out of the tensors A and B of different kinds,

$$T_{\mu\nu\sigma} = A_{\mu\nu}B_\sigma,$$
$$T^{\mu\nu\sigma\tau} = A^{\mu\nu}B^{\sigma\tau},$$
$$T^{\sigma\tau}_{\mu\nu} = A_{\mu\nu}B^{\sigma\tau}.$$

The proof of the tensor character of T is given directly by the representations (8), (10), (12), or by the laws of transformation (9), (11), (13). The equations (8), (10), (12) are themselves examples of outer multiplication of tensors of the first rank.

"*Contraction*" *of a Mixed Tensor.*—From any mixed tensor we may form a tensor whose rank is less by two, by equating an index of covariant with one of contravariant character, and summing with respect to this index ("contraction"). Thus, for example, from the mixed tensor of the fourth rank $A^{\sigma\tau}_{\mu\nu}$, we obtain the mixed tensor of the second rank,

$$A^\tau_\nu = A^{\mu\tau}_{\mu\nu} \quad \left(= \sum_\mu A^{\mu\tau}_{\mu\nu}\right),$$

and from this, by a second contraction, the tensor of zero rank,

$$A = A^\nu_\nu = A^{\mu\nu}_{\mu\nu}.$$

The proof that the result of contraction really possesses the tensor character is given either by the representation of a tensor according to the generalization of (12) in combination with (6), or by the generalization of (13).

Inner and Mixed Multiplication of Tensors.—These consist in a combination of outer multiplication with contraction.

Examples.—From the covariant tensor of the second rank $A_{\mu\nu}$ and the contravariant tensor of the first rank B^σ we form by outer multiplication the mixed tensor

$$D^\sigma_{\mu\nu} = A_{\mu\nu}B^\sigma.$$

On contraction with respect to the indices ν and σ, we obtain the covariant four-vector

$$D_\mu = D^\nu_{\mu\nu} = A_{\mu\nu}B^\nu.$$

This we call the inner product of the tensors $A_{\mu\nu}$ and B^σ. Analogously we form from the tensors $A_{\mu\nu}$ and $B^{\sigma\tau}$, by outer multiplication and double contraction, the inner product $A_{\mu\nu}B^{\mu\nu}$. By outer multiplication and one contraction, we obtain from $A_{\mu\nu}$ and $B^{\sigma\tau}$ the mixed tensor of the second rank $D^\tau_\mu = A_{\mu\nu}B^{\nu\tau}$. This operation may be aptly characterized as a mixed one, being "outer" with respect to the indices μ and τ, and "inner" with respect to the indices ν and σ.

We now prove a proposition which is often useful as evidence of tensor character. From what has just been explained, $A_{\mu\nu}B^{\mu\nu}$ is a scalar if $A_{\mu\nu}$ and $B^{\sigma\tau}$ are tensors. But we may also make the following assertion: If $A_{\mu\nu}B^{\mu\nu}$ is a scalar *for any choice of the tensor* $B^{\mu\nu}$, then $A_{\mu\nu}$ has tensor character. For, by hypothesis, for any substitution,

$$A'_{\sigma\tau}B'^{\sigma\tau} = A_{\mu\nu}B^{\mu\nu}.$$

But by an inversion of (9)

$$B^{\mu\nu} = \frac{\partial x_\mu}{\partial x'_\sigma}\frac{\partial x_\nu}{\partial x'_\tau}B'^{\sigma\tau}.$$

This, inserted in the above equation, gives

$$\left(A'_{\sigma\tau} - \frac{\partial x_\mu}{\partial x'_\sigma}\frac{\partial x_\nu}{\partial x'_\tau}A_{\mu\nu}\right)B'^{\sigma\tau} = 0.$$

This can only be satisfied for arbitrary values of $B'^{\sigma\tau}$ if the bracket vanishes. The result then follows by equation (11). This rule applies correspondingly to tensors of any rank and character, and the proof is analogous in all cases.

The rule may also be demonstrated in this form: If B^μ and C^ν are any vectors, and if, for all values of these, the inner product $A_{\mu\nu}B^\mu C^\nu$ is a scalar, then $A_{\mu\nu}$ is a covariant tensor. This latter proposition also holds good even if only the more special assertion is correct, that with any choice of the four-vector B^μ the inner product $A_{\mu\nu}B^\mu B^\nu$ is a scalar, if in addition it is known that $A_{\mu\nu}$ satisfies the condition of symmetry $A_{\mu\nu} = A_{\nu\mu}$. For by the method given above we prove the tensor character of $(A_{\mu\nu} + A_{\nu\mu})$, and from this the tensor character of $A_{\mu\nu}$ follows on account of symmetry. This also can be easily generalized to the case of covariant and contravariant tensors of any rank.

Finally, there follows from what has been proved, this law, which may also be generalized for any tensors: If for any choice of the four-vector B^ν the quantities $A_{\mu\nu}B^\nu$ form a tensor of the first rank, then $A_{\mu\nu}$ is a tensor of the second rank. For, if C^μ is any four-vector, then on account of the tensor character of $A_{\mu\nu}B^\nu$, the inner product $A_{\mu\nu}B^\nu C^\mu$ is a scalar for any choice of the two four-vectors B^ν and C^μ. From which the proposition follows.

§ 8. Some Aspects of the Fundamental Tensor $g_{\mu\nu}$

The Covariant Fundamental Tensor.—In the invariant expression for the square of the linear element,

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu,$$

the part played by the dx_μ is that of a contravariant vector which may be chosen at will. Since further, $g_{\mu\nu} = g_{\nu\mu}$, it follows from the considerations of the preceding paragraph that $g_{\mu\nu}$ is a covariant tensor of the second rank. We call it the "fundamental tensor." In what follows we deduce some properties of this tensor which, it is true, apply to any tensor of the second rank. But as the fundamental tensor plays a special part in our theory, which has its physical basis in the peculiar effects of gravitation, it so happens that the relations to be developed are of importance to us only in the case of the fundamental tensor.

The Contravariant Fundamental Tensor.—If in the determinant formed by the elements $g_{\mu\nu}$, we take the co-factor of each of the $g_{\mu\nu}$ and divide it by the determinant $g = | g_{\mu\nu} |$, we obtain certain quantities $g^{\mu\nu} (= g^{\nu\mu})$ which, as we shall demonstrate, form a contravariant tensor.

By a known property of determinants

$$g_{\mu\sigma}g^{\nu\sigma} = \delta_\mu^\nu \quad . \quad . \quad . \quad . \quad (16)$$

where the symbol δ_μ^ν denotes 1 or 0, according as $\mu = \nu$ or $\mu \neq \nu$.

Instead of the above expression for ds^2 we may thus write

$$g_{\mu\sigma}\delta_\nu^\sigma dx_\mu dx_\nu$$

or, by (16)

$$g_{\mu\sigma}g_{\nu\tau}g^{\sigma\tau}dx_\mu dx_\nu.$$

But, by the multiplication rules of the preceding paragraphs, the quantities

$$d\xi_\sigma = g_{\mu\sigma}dx_\mu$$

form a covariant four-vector, and in fact an arbitrary vector, since the dx_μ are arbitrary. By introducing this into our expression we obtain

$$ds^2 = g^{\sigma\tau}d\xi_\sigma d\xi_\tau.$$

Since this, with the arbitrary choice of the vector $d\xi_\sigma$, is a scalar, and $g^{\sigma\tau}$ by its definition is symmetrical in the indices σ and τ, it follows from the results of the preceding paragraph that $g^{\sigma\tau}$ is a contravariant tensor.

It further follows from (16) that δ_μ is also a tensor, which we may call the mixed fundamental tensor.

The Determinant of the Fundamental Tensor.—By the rule for the multiplication of determinants

$$| g_{\mu\alpha}g^{\alpha\nu} | = | g_{\mu\alpha} | \times | g^{\alpha\nu} |.$$

On the other hand

$$| g_{\mu\alpha}g^{\alpha\nu} | = | \delta_\mu^\nu | = 1.$$

It therefore follows that

$$| g_{\mu\nu} | \times | g^{\mu\nu} | = 1 \quad . \quad . \quad . \quad (17)$$

The Volume Scalar.—We seek first the law of transformation of the determinant $g = | g_{\mu\nu} |$. In accordance with (11)

$$g' = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \frac{\partial x}{\partial x'_\tau} g_{\mu\nu} \right|.$$

Hence, by a double application of the rule for the multiplication of determinants, it follows that

$$g' = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \right| \cdot \left| \frac{\partial x_\nu}{\partial x'_\tau} \right| \cdot | g_{\mu\nu} | = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \right|^2 g,$$

or

$$\sqrt{g'} = \left| \frac{\partial x_\mu}{\partial x'_\sigma} \right| \sqrt{g}.$$

On the other hand, the law of transformation of the element of volume

$$d\tau = \int dx_1 dx_2 dx_3 dx_4$$

is, in accordance with the theorem of Jacobi,

$$d\tau' = \left| \frac{\partial x'_\sigma}{\partial x_\mu} \right| d\tau.$$

By multiplication of the last two equations, we obtain

$$\sqrt{g'}d\tau' = \sqrt{g}d\tau \quad . \quad . \quad . \quad (18).$$

Instead of $\sqrt{g}$, we introduce in what follows the quantity $\sqrt{-g}$, which is always real on account of the hyperbolic character of the space-time continuum. The invariant $\sqrt{-g}d\tau$ is equal to the magnitude of the four-dimensional element of volume in the "local" system of reference, as measured with rigid rods and clocks in the sense of the special theory of relativity.

Note on the Character of the Space-time Continuum.—Our assumption that the special theory of relativity can always be applied to an infinitely small region, implies that ds^2 can always be expressed in accordance with (1) by means of real quantities $dX_1 \ldots dX_4$. If we denote by $d\tau_0$ the "natural" element of volume dX_1, dX_2, dX_3, dX_4, then

$$d\tau_0 = \sqrt{-g}d\tau \quad . \quad . \quad . \quad (18a)$$

If $\sqrt{-g}$ were to vanish at a point of the four-dimensional continuum, it would mean that at this point an infinitely small "natural" volume would correspond to a finite volume in the co-ordinates. Let us assume that this is never the case. Then g cannot change sign. We will assume that, in the sense of the special theory of relativity, g always has a finite negative value. This is a hypothesis as to the physical nature of the continuum under consideration, and at the same time a convention as to the choice of co-ordinates.

But if $-g$ is always finite and positive, it is natural to settle the choice of co-ordinates *a posteriori* in such a way that this quantity is always equal to unity. We shall see later that by such a restriction of the choice of co-ordinates it is possible to achieve an important simplification of the laws of nature.

In place of (18), we then have simply $d\tau' = d\tau$, from which, in view of Jacobi's theorem, it follows that

$$\left| \frac{\partial x'_\sigma}{\partial x_\mu} \right| = 1 \quad . \quad . \quad . \quad . \quad (19)$$

Thus, with this choice of co-ordinates, only substitutions for which the determinant is unity are permissible.

But it would be erroneous to believe that this step indicates a partial abandonment of the general postulate of relativity. We do not ask "What are the laws of nature which are co-variant in face of all substitutions for which the determinant is unity?" but our question is "What are the generally co-variant laws of nature?" It is not until we have formulated these that we simplify their expression by a particular choice of the system of reference.

The Formation of New Tensors by Means of the Fundamental Tensor.—Inner, outer, and mixed multiplication of a tensor by the fundamental tensor give tensors of different character and rank. For example,

$$A^\mu = g^{\mu\sigma} A_\sigma,$$
$$A = g_{\mu\nu} A^{\mu\nu}.$$

The following forms may be specially noted:—

$$A^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} A_{\alpha\beta},$$
$$A_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} A^{\alpha\beta}$$

(the "complements" of covariant and contravariant tensors respectively), and

$$B_{\mu\nu} = g_{\mu\nu} g^{\alpha\beta} A_{\alpha\beta}.$$

We call $B_{\mu\nu}$ the reduced tensor associated with $A_{\mu\nu}$. Similarly,

$$B^{\mu\nu} = g^{\mu\nu} g_{\alpha\beta} A^{\alpha\beta}.$$

It may be noted that $g^{\mu\nu}$ is nothing more than the complement of $g_{\mu\nu}$, since

$$g^{\mu\alpha} g^{\nu\beta} g_{\alpha\beta} = g^{\mu\alpha} \delta^\nu_\alpha = g^{\mu\nu}.$$

§ 9. The Equation of the Geodetic Line. The Motion of a Particle

As the linear element ds is defined independently of the system of co-ordinates, the line drawn between two points P and P′ of the four-dimensional continuum in such a way that $\int ds$ is stationary—a geodetic line—has a meaning which also is independent of the choice of co-ordinates. Its equation is

$$\delta \int_P^{P'} ds = 0 \quad . \quad . \quad . \quad . \quad (20)$$

Carrying out the variation in the usual way, we obtain from this equation four differential equations which define the geodetic line; this operation will be inserted here for the sake of completeness. Let λ be a function of the co-ordinates x_ν, and let this define a family of surfaces which intersect the required geodetic line as well as all the lines in immediate proximity to it which are drawn through the points P and P′. Any such line may then be supposed to be given by expressing its co-ordinates x_ν as functions of λ. Let the symbol δ indicate the transition from a point of the required geodetic to the point corresponding to the same λ on a neighbouring line. Then for (20) we may substitute

$$\left.\begin{aligned} \int_{\lambda_1}^{\lambda_2} \delta w \, d\lambda &= 0 \\ w^2 &= g_{\mu\nu} \frac{dx_\mu}{d\lambda} \frac{dx_\nu}{d\lambda} \end{aligned}\right\} \quad . \quad . \quad . \quad (20a)$$

But since

$$\delta w = \frac{1}{w}\left\{\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{d\lambda}\frac{dx_\nu}{d\lambda}\delta x_\sigma + g_{\mu\nu}\frac{dx_\mu}{d\lambda}\delta\left(\frac{dx_\nu}{d\lambda}\right)\right\},$$

and

$$\delta\left(\frac{dx_\nu}{d\lambda}\right) = \frac{d}{d\lambda}(\delta x_\nu),$$

we obtain from (20a), after a partial integration,

$$\int_{\lambda_1}^{\lambda_2}\kappa_\sigma\delta x_\sigma d\lambda = 0,$$

where

[14] $$\kappa_\sigma = \frac{d}{d\lambda}\left\{\frac{g_{\mu\nu}}{w}\frac{dx_\mu}{d\lambda}\right\} - \frac{1}{2w}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{d\lambda}\frac{dx_\nu}{d\lambda} \quad . \quad (20b)$$

Since the values of δx_σ are arbitrary, it follows from this that

$$\kappa_\sigma = 0 \quad . \quad . \quad . \quad . \quad . \quad (20c)$$

are the equations of the geodetic line.

If ds does not vanish along the geodetic line we may choose the "length of the arc" s, measured along the geodetic line, for the parameter λ. Then $w = 1$, and in place of (20c) we obtain

[15] $$g_{\mu\nu}\frac{d^2x_\mu}{ds^2} + \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\sigma}{ds}\frac{dx_\mu}{ds} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} = 0$$

or, by a mere change of notation,

$$g_{\alpha\sigma}\frac{d^2x_\alpha}{ds^2} + [\mu\nu, \sigma]\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} = 0 \quad . \quad . \quad (20d)$$

where, following Christoffel, we have written

$$[\mu\nu, \sigma] = \frac{1}{2}\left(\frac{\partial g_{\mu\sigma}}{\partial x_\nu} + \frac{\partial g_{\nu\sigma}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\sigma}\right) \quad . \quad . \quad (21)$$

Finally, if we multiply (20d) by $g^{\sigma\tau}$ (outer multiplication with respect to τ, inner with respect to σ), we obtain the equations of the geodetic line in the form

$$\frac{d^2x_\tau}{ds^2} + \{\mu\nu, \tau\}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} = 0 \quad . \quad . \quad . \quad (22)$$

where, following Christoffel, we have set

$$\{\mu\nu, \tau\} = g^{\tau\alpha}[\mu\nu, \alpha] \quad . \quad . \quad . \quad (23)$$

§ 10. The Formation of Tensors by Differentiation

With the help of the equation of the geodetic line we can now easily deduce the laws by which new tensors can be formed from old by differentiation. By this means we are able for the first time to formulate generally covariant differential equations. We reach this goal by repeated application of the following simple law :—

If in our continuum a curve is given, the points of which are specified by the arcual distance s measured from a fixed point on the curve, and if, further, ϕ is an invariant function of space, then $d\phi/ds$ is also an invariant. The proof lies in this, that ds is an invariant as well as $d\phi$.

As

$$\frac{d\phi}{ds} = \frac{\partial\phi}{\partial x_\mu}\frac{dx_\mu}{ds}$$

therefore

$$\psi = \frac{\partial\phi}{\partial x_\mu}\frac{dx_\mu}{ds}$$

is also an invariant, and an invariant for all curves starting from a point of the continuum, that is, for any choice of the vector dx_μ. Hence it immediately follows that

$$A_\mu = \frac{\partial\phi}{\partial x_\mu} \quad . \quad . \quad . \quad . \quad (24)$$

is a covariant four-vector—the "gradient" of ϕ.

According to our rule, the differential quotient

$$\chi = \frac{d\psi}{ds}$$

taken on a curve, is similarly an invariant. Inserting the value of ψ, we obtain in the first place

$$\chi = \frac{\partial^2\phi}{\partial x_\mu \partial x_\nu}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} + \frac{\partial\phi}{\partial x_\mu}\frac{d^2x_\mu}{ds^2}.$$

The existence of a tensor cannot be deduced from this forthwith. But if we may take the curve along which we have differentiated to be a geodetic, we obtain on substitution for d^2x_ν/ds^2 from (22),

$$\chi = \left(\frac{\partial^2\phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\}\frac{\partial\phi}{\partial x_\tau}\right)\frac{dx_\mu}{ds}\frac{dx_\nu}{ds}.$$

Since we may interchange the order of the differentiations, and since by (23) and (21) $\{\mu\nu, \tau\}$ is symmetrical in μ and ν, it follows that the expression in brackets is symmetrical in μ and ν. Since a geodetic line can be drawn in any direction from a point of the continuum, and therefore dx_μ/ds is a four-vector with the ratio of its components arbitrary, it follows from the results of § 7 that

$$A_{\mu\nu} = \frac{\partial^2\phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\}\frac{\partial\phi}{\partial x_\tau} \quad . \quad . \quad . \quad (25)$$

is a covariant tensor of the second rank. We have therefore come to this result: from the covariant tensor of the first rank

$$A_\mu = \frac{\partial\phi}{\partial x_\mu}$$

we can, by differentiation, form a covariant tensor of the second rank

$$A_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \{\mu\nu, \tau\}A_\tau \quad . \quad . \quad . \quad (26)$$

We call the tensor $A_{\mu\nu}$ the " extension " (covariant derivative) of the tensor A_μ. In the first place we can readily show that the operation leads to a tensor, even if the vector A_μ cannot be represented as a gradient. To see this, we first observe that

$$\psi\frac{\partial\phi}{\partial x_\mu}$$

is a covariant vector, if ψ and ϕ are scalars. The sum of four such terms

$$S_\mu = \psi^{(1)}\frac{\partial\phi^{(1)}}{\partial x_\mu} + . + . + \psi^{(4)}\frac{\partial\phi^{(4)}}{\partial x_\mu},$$

is also a covariant vector, if $\psi^{(1)}, \phi^{(1)} \ldots \psi^{(4)}, \phi^{(4)}$ are scalars. But it is clear that any covariant vector can be represented in the form S_μ. For, if A_μ is a vector whose components are any given functions of the x_ν, we have only to put (in terms of the selected system of co-ordinates)

$$\begin{aligned} \psi^{(1)} &= A_1, & \phi^{(1)} &= x_1, \\ \psi^{(2)} &= A_2, & \phi^{(2)} &= x_2, \\ \psi^{(3)} &= A_3, & \phi^{(3)} &= x_3, \\ \psi^{(4)} &= A_4, & \phi^{(4)} &= x_4, \end{aligned}$$

in order to ensure that S_μ shall be equal to A_μ.

Therefore, in order to demonstrate that $A_{\mu\nu}$ is a tensor if *any* covariant vector is inserted on the right-hand side for A_μ, we only need show that this is so for the vector S_μ. But for this latter purpose it is sufficient, as a glance at the right-hand side of (26) teaches us, to furnish the proof for the case

$$A_\mu = \psi\frac{\partial\phi}{\partial x_\mu}.$$

Now the right-hand side of (25) multiplied by ψ,

$$\psi\frac{\partial^2\phi}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \tau\}\psi\frac{\partial\phi}{\partial x_\tau}$$

is a tensor. Similarly

$$\frac{\partial\psi}{\partial x_\mu}\frac{\partial\phi}{\partial x_\nu}$$

being the outer product of two vectors, is a tensor. By addition, there follows the tensor character of

$$\frac{\partial}{\partial x_\nu}\left(\psi\frac{\partial\phi}{\partial x_\mu}\right) - \{\mu\nu, \tau\}\left(\psi\frac{\partial\phi}{\partial x_\tau}\right).$$

As a glance at (26) will show, this completes the demonstration for the vector

$$\psi\frac{\partial\phi}{\partial x_\mu}$$

and consequently, from what has already been proved, for any vector A_μ.

By means of the extension of the vector, we may easily define the " extension " of a covariant tensor of any rank. This operation is a generalization of the extension of a vector. We restrict ourselves to the case of a tensor of the second rank, since this suffices to give a clear idea of the law of formation.

As has already been observed, any covariant tensor of the second rank can be represented * as the sum of tensors of the

* By outer multiplication of the vector with arbitrary components $A_{11}, A_{12}, A_{13}, A_{14}$ by the vector with components 1, 0, 0, 0, we produce a tensor with components

$$\begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0. \end{matrix}$$

By the addition of four tensors of this type, we obtain the tensor $A_{\mu\nu}$ with any ssigned components.

type $A_\mu B_\nu$. It will therefore be sufficient to deduce the expression for the extension of a tensor of this special type. By (26) the expressions

$$\frac{\partial A_\mu}{\partial x_\sigma} - \{\sigma\mu, \tau\}A_\tau,$$

$$\frac{\partial B_\nu}{\partial x_\sigma} - \{\sigma\nu, \tau\}B_\tau,$$

are tensors. On outer multiplication of the first by B_ν, and of the second by A_μ, we obtain in each case a tensor of the third rank. By adding these, we have the tensor of the third rank

$$A_{\mu\nu\sigma} = \frac{\partial A_{\mu\nu}}{\partial x_\sigma} - \{\sigma\mu, \tau\}A_{\tau\nu} - \{\sigma\nu, \tau\}A_{\mu\tau} \quad . \quad . \quad (27)$$

where we have put $A_{\mu\nu} = A_\mu B_\nu$. As the right-hand side of (27) is linear and homogeneous in the $A_{\mu\nu}$ and their first derivatives, this law of formation leads to a tensor, not only in the case of a tensor of the type $A_\mu B_\nu$, but also in the case of a sum of such tensors, i.e. in the case of any covariant tensor of the second rank. We call $A_{\mu\nu\sigma}$ the extension of the tensor $A_{\mu\nu}$.

It is clear that (26) and (24) concern only special cases of extension (the extension of the tensors of rank one and zero respectively).

In general, all special laws of formation of tensors are included in (27) in combination with the multiplication of tensors.

§ 11. Some Cases of Special Importance

The Fundamental Tensor.—We will first prove some lemmas which will be useful hereafter. By the rule for the differentiation of determinants

$$dg = g^{\mu\nu}dg_{\mu\nu} = -g_{\mu\nu}dg^{\mu\nu} \quad . \quad . \quad (28)$$

The last member is obtained from the last but one, if we bear in mind that $g_{\mu\nu}g^{\mu'\nu} = \delta_\mu^{\mu'}$, so that $g_{\mu\nu}g^{\mu\nu} = 4$, and consequently

$$g_{\mu\nu}dg^{\mu\nu} + g^{\mu\nu}dg_{\mu\nu} = 0.$$

From (28), it follows that

$$\frac{1}{\sqrt{-g}}\frac{\partial\sqrt{-g}}{\partial x_\sigma} = \tfrac{1}{2}\frac{\partial \log(-g)}{\partial x_\sigma} = \tfrac{1}{2}g^{\mu\nu}\frac{\partial g_{\mu\nu}}{\partial x_\sigma} = \tfrac{1}{2}g_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}. \quad (29)$$

Further, from $g_{\mu\sigma}g^{\nu\sigma} = \delta_\mu^\nu$, it follows on differentiation that

$$\left.\begin{aligned} g_{\mu\sigma}dg^{\nu\sigma} &= -g^{\nu\sigma}dg_{\mu\sigma} \\ g_{\mu\sigma}\frac{\partial g^{\nu\sigma}}{\partial x_\lambda} &= -g^{\nu\sigma}\frac{\partial g_{\mu\sigma}}{\partial x_\lambda} \end{aligned}\right\} \quad . \quad . \quad . \quad (30)$$

From these, by mixed multiplication by $g^{\sigma\tau}$ and $g_{\nu\lambda}$ respectively, and a change of notation for the indices, we have

$$\left.\begin{aligned} dg^{\mu\nu} &= -g^{\mu\alpha}g^{\nu\beta}\,dg_{\alpha\beta} \\ \frac{\partial g^{\mu\nu}}{\partial x_\sigma} &= -g^{\mu\alpha}g^{\nu\beta}\frac{\partial g_{\alpha\beta}}{\partial x_\sigma} \end{aligned}\right\} \quad . \quad . \quad . \quad (31)$$

and

$$\left.\begin{aligned} dg_{\mu\nu} &= -g_{\mu\alpha}g_{\nu\beta}\,dg^{\alpha\beta} \\ \frac{\partial g_{\mu\nu}}{\partial x_\sigma} &= -g_{\mu\alpha}g_{\nu\beta}\frac{\partial g^{\alpha\beta}}{\partial x_\sigma} \end{aligned}\right\} \quad . \quad . \quad . \quad (32)$$

The relation (31) admits of a transformation, of which we also have frequently to make use. From (21)

$$\frac{\partial g_{\alpha\beta}}{\partial x_\sigma} = [\alpha\sigma, \beta] + [\beta\sigma, \alpha] \quad . \quad . \quad . \quad (33)$$

Inserting this in the second formula of (31), we obtain, in view of (23)

$$\frac{\partial g^{\mu\nu}}{\partial x_\sigma} = -g^{\mu\tau}\{\tau\sigma, \nu\} - g^{\nu\tau}\{\tau\sigma, \mu\} \quad . \quad . \quad (34)$$ [16]

Substituting the right-hand side of (34) in (29), we have

$$\frac{1}{\sqrt{-g}}\frac{\partial\sqrt{-g}}{\partial x_\sigma} = \{\mu\sigma, \mu\} \quad . \quad . \quad (29a)$$ [17]

The "Divergence" of a Contravariant Vector.—If we take the inner product of (26) by the contravariant fundamental tensor $g^{\mu\nu}$, the right-hand side, after a transformation of the first term, assumes the form

$$\frac{\partial}{\partial x_\nu}(g^{\mu\nu}A_\mu) - A_\mu\frac{\partial g^{\mu\nu}}{\partial x_\nu} - \tfrac{1}{2}g^{\tau\alpha}\left(\frac{\partial g_{\mu\alpha}}{\partial x_\nu} + \frac{\partial g_{\nu\alpha}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right)g^{\mu\nu}A_\tau.$$ [18]

[19] In accordance with (31) and (29), the last term of this expression may be written

[20]
$$\tfrac{1}{2}\frac{\partial g^{\tau\nu}}{\partial x_\nu}A_\tau + \tfrac{1}{2}\frac{\partial g^{\tau\mu}}{\partial x_\mu}A_\tau + \frac{1}{\sqrt{-g}}\frac{\partial\sqrt{-g}}{\partial x_\alpha}g^{\mu\nu}A_\tau.$$

As the symbols of the indices of summation are immaterial, the first two terms of this expression cancel the second of the one above. If we then write $g^{\mu\nu}A_\mu = A^\nu$, so that A^ν like A_μ is an arbitrary vector, we finally obtain

$$\Phi = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x_\nu}(\sqrt{-g}A^\nu). \quad . \quad . \quad (35)$$

This scalar is the *divergence* of the contravariant vector A^ν.

The "Curl" of a Covariant Vector.—The second term in (26) is symmetrical in the indices μ and ν. Therefore $A_{\mu\nu} - A_{\nu\mu}$ is a particularly simply constructed antisymmetrical tensor. We obtain

$$B_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu} \quad . \quad . \quad . \quad (36)$$

Antisymmetrical Extension of a Six-vector.—Applying (27) to an antisymmetrical tensor of the second rank $A_{\mu\nu}$, forming in addition the two equations which arise through cyclic permutations of the indices, and adding these three equations, we obtain the tensor of the third rank

$$B_{\mu\nu\sigma} = A_{\mu\nu\sigma} + A_{\nu\sigma\mu} + A_{\sigma\mu\nu} = \frac{\partial A_{\mu\nu}}{\partial x_\sigma} + \frac{\partial A_{\nu\sigma}}{\partial x_\mu} + \frac{\partial A_{\sigma\mu}}{\partial x_\nu} \quad (37)$$

which it is easy to prove is antisymmetrical.

The Divergence of a Six-vector.—Taking the mixed product of (27) by $g^{\mu\alpha}g^{\nu\beta}$, we also obtain a tensor. The first term on the right-hand side of (27) may be written in the form

$$\frac{\partial}{\partial x_\sigma}(g^{\mu\alpha}g^{\nu\beta}A_{\mu\nu}) - g^{\mu\alpha}\frac{\partial g^{\nu\beta}}{\partial x_\sigma}A_{\mu\nu} - g^{\nu\beta}\frac{\partial g^{\mu\alpha}}{\partial x_\sigma}A_{\mu\nu}.$$

If we write $A^{\alpha\beta}_\sigma$ for $g^{\mu\alpha}g^{\nu\beta}A_{\mu\nu\sigma}$ and $A^{\alpha\beta}$ for $g^{\mu\alpha}g^{\nu\beta}A_{\mu\nu}$, and in the transformed first term replace

$$\frac{\partial g^{\nu\beta}}{\partial x_\sigma} \text{ and } \frac{\partial g^{\mu\alpha}}{\partial x_\sigma}$$

by their values as given by (34), there results from the right-hand side of (27) an expression consisting of seven terms, of which four cancel, and there remains

$$A^{\alpha\beta}_\sigma = \frac{\partial A^{\alpha\beta}}{\partial x_\sigma} + \{\sigma\gamma, \alpha\}A^{\gamma\beta} + \{\sigma\gamma, \beta\}A^{\alpha\gamma}. \quad . \quad (38)$$

This is the expression for the extension of a contravariant tensor of the second rank, and corresponding expressions for the extension of contravariant tensors of higher and lower rank may also be formed.

We note that in an analogous way we may also form the extension of a mixed tensor:—

$$A^\alpha_{\mu\sigma} = \frac{\partial A^\alpha_\mu}{\partial x_\sigma} - \{\sigma\mu, \tau\}A^\alpha_\tau + \{\sigma\tau, \alpha\}A^\tau_\mu. \quad . \quad (39)$$

On contracting (38) with respect to the indices β and σ (inner multiplication by δ^σ_β), we obtain the vector

$$A^\alpha = \frac{\partial A^{\alpha\beta}}{\partial x_\beta} + \{\beta\gamma, \beta\}A^{\alpha\gamma} + \{\beta\gamma, \alpha\}A^{\gamma\beta}.$$

On account of the symmetry of $\{\beta\gamma, \alpha\}$ with respect to the indices β and γ, the third term on the right-hand side vanishes, if $A^{\alpha\beta}$ is, as we will assume, an antisymmetrical tensor. The second term allows itself to be transformed in accordance with (29a). Thus we obtain

$$A^\alpha = \frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}A^{\alpha\beta})}{\partial x_\beta} \quad . \quad . \quad . \quad (40)$$

This is the expression for the divergence of a contravariant six-vector.

The Divergence of a Mixed Tensor of the Second Rank.—Contracting (39) with respect to the indices α and σ, and taking (29a) into consideration, we obtain

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A^\sigma_\mu)}{\partial x_\sigma} - \{\sigma\mu, \tau\}\sqrt{-g}A^\sigma_\tau. \quad (41)$$

If we introduce the contravariant tensor $A^{\rho\sigma} = g^{\rho\tau}A^\sigma_\tau$ in the last term, it assumes the form

$$-[\sigma\mu, \rho]\sqrt{-g}A^{\rho\sigma}.$$

If, further, the tensor $A^{\rho\sigma}$ is symmetrical, this reduces to

$$-\tfrac{1}{2}\sqrt{-g}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}A^{\rho\sigma}.$$

Had we introduced, instead of $A^{\rho\sigma}$, the covariant tensor $A_{\rho\sigma} = g_{\rho\alpha}g_{\sigma\beta}A^{\alpha\beta}$, which is also symmetrical, the last term, by virtue of (31), would assume the form

$$\tfrac{1}{2}\sqrt{-g}\frac{\partial g^{\rho\sigma}}{\partial x_\mu}A_{\rho\sigma}.$$

In the case of symmetry in question, (41) may therefore be replaced by the two forms

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A_\mu^\sigma)}{\partial x_\sigma} - \tfrac{1}{2}\frac{\partial g_{\rho\sigma}}{\partial x_\mu}\sqrt{-g}A^{\rho\sigma} \quad . \quad (41a)$$

$$\sqrt{-g}A_\mu = \frac{\partial(\sqrt{-g}A_\mu^\sigma)}{\partial x_\sigma} + \tfrac{1}{2}\frac{\partial g^{\rho\sigma}}{\partial x_\mu}\sqrt{-g}A_{\rho\sigma} \quad . \quad (41b)$$

which we have to employ later on.

§ 12. The Riemann-Christoffel Tensor

We now seek the tensor which can be obtained from the fundamental tensor *alone*, by differentiation. At first sight the solution seems obvious. We place the fundamental tensor of the $g_{\mu\nu}$ in (27) instead of any given tensor $A_{\mu\nu}$, and thus have a new tensor, namely, the extension of the fundamental tensor. But we easily convince ourselves that this extension vanishes identically. We reach our goal, however, in the following way. In (27) place

$$A_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \{\mu\nu, \rho\}A_\rho,$$

i.e. the extension of the four-vector A_μ. Then (with a somewhat different naming of the indices) we get the tensor of the third rank

$$A_{\mu\sigma\tau} = \frac{\partial^2 A_\mu}{\partial x_\sigma \partial x_\tau} - \{\mu\sigma, \rho\}\frac{\partial A_\rho}{\partial x_\tau} - \{\mu\tau, \rho\}\frac{\partial A_\rho}{\partial x_\sigma} - \{\sigma\tau, \rho\}\frac{\partial A_\mu}{\partial x_\rho}$$
$$+ \left[-\frac{\partial}{\partial x_\tau}\{\mu\sigma, \rho\} + \{\mu\tau, \alpha\}\{\alpha\sigma, \rho\} + \{\sigma\tau, \alpha\}\{\alpha\mu, \rho\}\right]A_\rho.$$

This expression suggests forming the tensor $A_{\mu\sigma\tau} - A_{\mu\tau\sigma}$. For, if we do so, the following terms of the expression for $A_{\mu\sigma\tau}$ cancel those of $A_{\mu\tau\sigma}$, the first, the fourth, and the member corresponding to the last term in square brackets; because all these are symmetrical in σ and τ. The same holds good for the sum of the second and third terms. Thus we obtain

$$A_{\mu\sigma\tau} - A_{\mu\tau\sigma} = B^\rho_{\mu\sigma\tau}A_\rho \quad . \quad . \quad . \quad (42)$$

where

$$B^\rho_{\mu\sigma\tau} = -\frac{\partial}{\partial x_\tau}\{\mu\sigma, \rho\} + \frac{\partial}{\partial x_\sigma}\{\mu\tau, \rho\} - \{\mu\sigma, \alpha\}\{\alpha\tau, \rho\}$$
$$+ \{\mu\tau, \alpha\}\{\alpha\sigma, \rho\} \quad (43)$$

The essential feature of the result is that on the right side of (42) the A_ρ occur alone, without their derivatives. From the tensor character of $A_{\mu\sigma\tau} - A_{\mu\tau\sigma}$ in conjunction with the fact that A_ρ is an arbitrary vector, it follows, by reason of § 7, that $B^\rho_{\mu\sigma\tau}$ is a tensor (the Riemann-Christoffel tensor).

The mathematical importance of this tensor is as follows: If the continuum is of such a nature that there is a co-ordinate system with reference to which the $g_{\mu\nu}$ are constants, then all the $B^\rho_{\mu\sigma\tau}$ vanish. If we choose any new system of co-ordinates in place of the original ones, the $g_{\mu\nu}$ referred thereto will not be constants, but in consequence of its tensor nature, the transformed components of $B^\rho_{\mu\sigma\tau}$ will still vanish in the new system. Thus the vanishing of the Riemann tensor is a necessary condition that, by an appropriate choice of the system of reference, the $g_{\mu\nu}$ may be constants. In our problem this corresponds to the case in which,* with a suitable choice of the system of reference, the special theory of relativity holds good for a *finite* region of the continuum. [21]

Contracting (43) with respect to the indices τ and ρ we obtain the covariant tensor of second rank

* The mathematicians have proved that this is also a *sufficient* condition.

$$\left.\begin{aligned} \mathrm{G}_{\mu\nu} &= \mathrm{B}^{\rho}_{\mu\nu\rho} = \mathrm{R}_{\mu\nu} + \mathrm{S}_{\mu\nu} \\ \text{where} \quad & \\ \mathrm{R}_{\mu\nu} &= -\frac{\partial}{\partial x_\alpha}\{\mu\nu, \alpha\} + \{\mu\alpha, \beta\}\{\nu\beta, \alpha\} \\ \mathrm{S}_{\mu\nu} &= \frac{\partial^2 \log\sqrt{-g}}{\partial x_\mu \partial x_\nu} - \{\mu\nu, \alpha\}\frac{\partial \log\sqrt{-g}}{\partial x_\alpha} \end{aligned}\right\} \qquad (44)$$

Note on the Choice of Co-ordinates.—It has already been observed in § 8, in connexion with equation (18a), that the choice of co-ordinates may with advantage be made so that $\sqrt{-g} = 1$. A glance at the equations obtained in the last two sections shows that by such a choice the laws of formation of tensors undergo an important simplification. This applies particularly to $\mathrm{G}_{\mu\nu}$, the tensor just developed, which plays a fundamental part in the theory to be set forth. For this specialization of the choice of co-ordinates brings about the vanishing of $\mathrm{S}_{\mu\nu}$, so that the tensor $\mathrm{G}_{\mu\nu}$ reduces to $\mathrm{R}_{\mu\nu}$.

On this account I shall hereafter give all relations in the simplified form which this specialization of the choice of co-ordinates brings with it. It will then be an easy matter to revert to the *generally* covariant equations, if this seems desirable in a special case.

C. Theory of the Gravitational Field

§ 13. Equations of Motion of a Material Point in the Gravitational Field. Expression for the Field-components of Gravitation

A freely movable body not subjected to external forces moves, according to the special theory of relativity, in a straight line and uniformly. This is also the case, according to the general theory of relativity, for a part of four-dimensional space in which the system of co-ordinates K_0, may be, and is, so chosen that they have the special constant values given in (4).

If we consider precisely this movement from any chosen system of co-ordinates K_1, the body, observed from K_1, moves, according to the considerations in § 2, in a gravitational field. The law of motion with respect to K_1 results without difficulty from the following consideration. With respect to K_0 the law of motion corresponds to a four-dimensional straight line, i.e. to a geodetic line. Now since the geodetic line is defined independently of the system of reference, its equations will also be the equation of motion of the material point with respect to K_1. If we set

$$\Gamma^{\tau}_{\mu\nu} = -\{\mu\nu, \tau\} \quad . \quad . \quad . \quad (45)$$

the equation of the motion of the point with respect to K_1, becomes

$$\frac{d^2 x_\tau}{ds^2} = \Gamma^{\tau}_{\mu\nu}\frac{dx_\mu}{ds}\frac{dx_\nu}{ds} \quad . \quad . \quad . \quad (46)$$

We now make the assumption, which readily suggests itself, that this covariant system of equations also defines the motion of the point in the gravitational field in the case when there is no system of reference K_0, with respect to which the special theory of relativity holds good in a finite region. We have all the more justification for this assumption as (46) contains only *first* derivatives of the $g_{\mu\nu}$, between which even in the special case of the existence of K_0, no relations subsist.*

If the $\Gamma^{\tau}_{\mu\nu}$ vanish, then the point moves uniformly in a straight line. These quantities therefore condition the deviation of the motion from uniformity. They are the components of the gravitational field.

§ 14. The Field Equations of Gravitation in the Absence of Matter

We make a distinction hereafter between "gravitational field" and "matter" in this way, that we denote everything but the gravitational field as "matter." Our use of the word therefore includes not only matter in the ordinary sense, but the electromagnetic field as well.

Our next task is to find the field equations of gravitation in the absence of matter. Here we again apply the method

* It is only between the second (and first) derivatives that, by § 12, the relations $\mathrm{B}^{\rho}_{\mu\sigma\tau} = 0$ subsist.

employed in the preceding paragraph in formulating the equations of motion of the material point. A special case in which the required equations must in any case be satisfied is that of the special theory of relativity, in which the $g_{\mu\nu}$ have certain constant values. Let this be the case in a certain finite space in relation to a definite system of co-ordinates K_0. Relatively to this system all the components of the Riemann tensor $B^{\rho}_{\mu\sigma\tau}$, defined in (43), vanish. For the space under consideration they then vanish, also in any other system of co-ordinates.

Thus the required equations of the matter-free gravitational field must in any case be satisfied if all $B^{\rho}_{\mu\sigma\tau}$ vanish. But this condition goes too far. For it is clear that, e.g., the gravitational field generated by a material point in its environment certainly cannot be "transformed away" by any choice of the system of co-ordinates, i.e. it cannot be transformed to the case of constant $g_{\mu\nu}$.

This prompts us to require for the matter-free gravitational field that the symmetrical tensor $G_{\mu\nu}$, derived from the tensor $B^{\rho}_{\mu\nu\tau}$, shall vanish. Thus we obtain ten equations for the ten quantities $g_{\mu\nu}$, which are satisfied in the special case of the vanishing of all $B^{\rho}_{\mu\nu\tau}$. With the choice which we have made of a system of co-ordinates, and taking (44) into consideration, the equations for the matter-free field are

$$\left.\begin{aligned}\frac{\partial\Gamma^{\alpha}_{\mu\nu}}{\partial x_\alpha} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} &= 0\\ \sqrt{-g} &= 1\end{aligned}\right\} \quad . \quad . \quad . \quad (47)$$

It must be pointed out that there is only a minimum of arbitrariness in the choice of these equations. For besides $G_{\mu\nu}$ there is no tensor of second rank which is formed from the $g_{\mu\nu}$ and its derivatives, contains no derivations higher than second, and is linear in these derivatives.*

These equations, which proceed, by the method of pure mathematics, from the requirement of the general theory of relativity, give us, in combination with the equations of motion (46), to a first approximation Newton's law of attraction, and to a second approximation the explanation of the motion of the perihelion of the planet Mercury discovered by Leverrier (as it remains after corrections for perturbation have been made). These facts must, in my opinion, be taken as a convincing proof of the correctness of the theory.

* Properly speaking, this can be affirmed only of the tensor

$$G_{\mu\nu} + \lambda g_{\mu\nu}g^{\alpha\beta}G_{\alpha\beta},$$

where λ is a constant. If, however, we set this tensor = 0, we come back again to the equations $G_{\mu\nu} = 0$.

§ 15. The Hamiltonian Function for the Gravitational Field. Laws of Momentum and Energy

To show that the field equations correspond to the laws of momentum and energy, it is most convenient to write them in the following Hamiltonian form :—

$$\left.\begin{aligned}\delta\int H d\tau &= 0\\ H &= g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\\ \sqrt{-g} &= 1\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (47a)$$

where, on the boundary of the finite four-dimensional region of integration which we have in view, the variations vanish.

We first have to show that the form (47a) is equivalent to the equations (47). For this purpose we regard H as a function of the $g^{\mu\nu}$ and the $g^{\mu\nu}_{\sigma}$ $(= \partial g^{\mu\nu}/\partial x_\sigma)$.

Then in the first place

$$\begin{aligned}\delta H &= \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\,\delta g^{\mu\nu} + 2g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\,\delta\Gamma^{\beta}_{\nu\alpha}\\ &= -\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\,\delta g^{\mu\nu} + 2\Gamma^{\alpha}_{\mu\beta}\,\delta(g^{\mu\nu}\Gamma^{\beta}_{\nu\alpha}).\end{aligned}$$

But

$$\delta\left(g^{\mu\nu}\Gamma^{\beta}_{\nu\alpha}\right) = -\tfrac{1}{2}\delta\left[g^{\mu\nu}g^{\beta\lambda}\left(\frac{\partial g_{\nu\lambda}}{\partial x_\alpha} + \frac{\partial g_{\alpha\lambda}}{\partial x_\nu} - \frac{\partial g_{\alpha\nu}}{\partial x_\lambda}\right)\right]. \quad [22]$$

The terms arising from the last two terms in round brackets are of different sign, and result from each other (since the denomination of the summation indices is immaterial) through interchange of the indices μ and β. They cancel each other in the expression for δH, because they are multiplied by the quantity $\Gamma^{\alpha}_{\mu\beta}$, which is symmetrical with respect to the indices μ and β. Thus there remains only the first term in round brackets to be considered, so that, taking (31) into account, we obtain

$$\delta H = -\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}\delta g^{\mu\nu} + \Gamma^{\alpha}_{\mu\beta}\delta g^{\mu\beta}_{\alpha}.$$

Thus

[23]

$$\left.\begin{aligned} \frac{\partial H}{\partial g^{\mu\nu}} &= -\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} \\ \frac{\partial H}{\partial g^{\mu\nu}_{\sigma}} &= \Gamma^{\sigma}_{\mu\nu} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (48)$$

Carrying out the variation in (47a), we get in the first place

$$\frac{\partial}{\partial x_{\alpha}}\left(\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial g^{\mu\nu}} = 0, \quad . \quad . \quad . \quad (47b)$$

which, on account of (48), agrees with (47), as was to be proved.

If we multiply (47b) by $g^{\mu\nu}_{\sigma}$, then because

$$\frac{\partial g^{\mu\nu}_{\sigma}}{\partial x_{\alpha}} = \frac{\partial g^{\mu\nu}_{\alpha}}{\partial x_{\sigma}}$$

and, consequently,

$$g^{\mu\nu}_{\sigma}\frac{\partial}{\partial x_{\alpha}}\left(\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) = \frac{\partial}{\partial x_{\alpha}}\left(g^{\mu\nu}_{\sigma}\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\frac{\partial g^{\mu\nu}_{\alpha}}{\partial x_{\sigma}},$$

we obtain the equation

$$\frac{\partial}{\partial x_{\alpha}}\left(g^{\mu\nu}_{\sigma}\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}}\right) - \frac{\partial H}{\partial x_{\sigma}} = 0$$

or *

$$\left.\begin{aligned} \frac{\partial t^{\alpha}_{\sigma}}{\partial x_{\alpha}} &= 0 \\ -2\kappa t^{\alpha}_{\sigma} &= g^{\mu\nu}_{\sigma}\frac{\partial H}{\partial g^{\mu\nu}_{\alpha}} - \delta^{\alpha}_{\sigma}H \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (49)$$

where, on account of (48), the second equation of (47), and (34)

[24]

$$\kappa t^{\alpha}_{\sigma} = \tfrac{1}{2}\delta^{\alpha}_{\sigma}g^{\mu\nu}\Gamma^{\lambda}_{\mu\beta}\Gamma^{\beta}_{\nu\lambda} - g^{\mu\nu}\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\sigma} \quad . \quad . \quad (50)$$

* The reason for the introduction of the factor -2κ will be apparent later.

It is to be noticed that t^{α}_{σ} is not a tensor; on the other hand (49) applies to all systems of co-ordinates for which $\sqrt{-g} = 1$. This equation expresses the law of conservation of momentum and of energy for the gravitational field. Actually the integration of this equation over a three-dimensional volume V yields the four equations

$$\frac{d}{dx_4}\int t^4_{\sigma}dV = \int(lt^1_{\sigma} + mt^2_{\sigma} + nt^3_{\sigma})dS. \quad . \quad (49a)$$

where l, m, n denote the direction-cosines of direction of the inward drawn normal at the element dS of the bounding surface (in the sense of Euclidean geometry). We recognize in this the expression of the laws of conservation in their usual form. The quantities t^{α}_{σ} we call the "energy components" of the gravitational field.

I will now give equations (47) in a third form, which is particularly useful for a vivid grasp of our subject. By multiplication of the field equations (47) by $g^{\nu\sigma}$ these are obtained in the "mixed" form. Note that

$$g^{\nu\sigma}\frac{\partial\Gamma^{\alpha}_{\mu\nu}}{\partial x_{\alpha}} = \frac{\partial}{\partial x_{\alpha}}\left(g^{\nu\sigma}\Gamma^{\alpha}_{\mu\nu}\right) - \frac{\partial g^{\nu\sigma}}{\partial x_{\alpha}}\Gamma^{\alpha}_{\mu\nu},$$

which quantity, by reason of (34), is equal to

$$\frac{\partial}{\partial x_{\alpha}}\left(g^{\nu\sigma}\Gamma^{\alpha}_{\mu\nu}\right) - g^{\nu\beta}\Gamma^{\sigma}_{\alpha\beta}\Gamma^{\alpha}_{\mu\nu} - g^{\sigma\beta}\Gamma^{\nu}_{\beta\alpha}\Gamma^{\alpha}_{\mu\nu},$$

or (with different symbols for the summation indices)

$$\frac{\partial}{\partial x_{\alpha}}\left(g^{\sigma\beta}\Gamma^{\alpha}_{\mu\beta}\right) - g^{\gamma\delta}\Gamma^{\sigma}_{\gamma\beta}\Gamma^{\beta}_{\delta\mu} - g^{\nu\sigma}\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha}.$$

The third term of this expression cancels with the one arising from the second term of the field equations (47); using relation (50), the second term may be written

$$\kappa(t^{\sigma}_{\mu} - \tfrac{1}{2}\delta^{\sigma}_{\mu}t),$$

where $t = t^{\alpha}_{\alpha}$. Thus instead of equations (47) we obtain

$$\left.\begin{aligned} \frac{\partial}{\partial x_{\alpha}}\left(g^{\sigma\beta}\Gamma^{\alpha}_{\mu\beta}\right) &= -\kappa(t^{\sigma}_{\mu} - \tfrac{1}{2}\delta^{\sigma}_{\mu}t) \\ \sqrt{-g} &= 1 \end{aligned}\right\} \quad . \quad . \quad (51)$$

§ 16. The General Form of the Field Equations of Gravitation

The field equations for matter-free space formulated in § 15 are to be compared with the field equation

$$\nabla^2\phi = 0$$

of Newton's theory. We require the equation corresponding to Poisson's equation

$$\nabla^2\phi = 4\pi\kappa\rho,$$

where ρ denotes the density of matter.

The special theory of relativity has led to the conclusion that inert mass is nothing more or less than energy, which finds its complete mathematical expression in a symmetrical tensor of second rank, the energy-tensor. Thus in the general theory of relativity we must introduce a corresponding energy-tensor of matter T^α_σ, which, like the energy-components t_σ [equations (49) and (50)] of the gravitational field, will have mixed character, but will pertain to a symmetrical covariant tensor.*

The system of equation (51) shows how this energy-tensor (corresponding to the density ρ in Poisson's equation) is to be introduced into the field equations of gravitation. For if we consider a complete system (e.g. the solar system), the total mass of the system, and therefore its total gravitating action as well, will depend on the total energy of the system, and therefore on the ponderable energy together with the gravitational energy. This will allow itself to be expressed by introducing into (51), in place of the energy-components of the gravitational field alone, the sums $t^\sigma_\mu + T^\sigma_\mu$ of the energy-components of matter and of gravitational field. Thus instead of (51) we obtain the tensor equation

$$\left.\begin{aligned}\frac{\partial}{\partial x_\alpha}(g^{\sigma\beta}\Gamma^\alpha_{\mu\beta}) &= -\kappa[(t^\sigma_\mu + T^\sigma_\mu) - \tfrac{1}{2}\delta^\sigma_\mu(t + T)],\\ \sqrt{-g} &= 1\end{aligned}\right\} \quad . \quad (52)$$

where we have set $T = T^\mu_\mu$ (Laue's scalar). These are the

[25] * $g_{\alpha\tau}T^\alpha_\sigma = T_{\sigma\tau}$ and $g^{\sigma\beta}T^\alpha_\sigma = T^{\alpha\beta}$ are to be symmetrical tensors.

required general field equations of gravitation in mixed form. Working back from these, we have in place of (47)

$$\left.\begin{aligned}\frac{\partial}{\partial x_\alpha}\Gamma^\alpha_{\mu\nu} + \Gamma^\alpha_{\mu\beta}\Gamma^\beta_{\nu\alpha} &= -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T),\\ \sqrt{-g} &= 1\end{aligned}\right\} \quad . \quad (53)$$

It must be admitted that this introduction of the energy-tensor of matter is not justified by the relativity postulate alone. For this reason we have here deduced it from the requirement that the energy of the gravitational field shall act gravitatively in the same way as any other kind of energy. But the strongest reason for the choice of these equations lies in their consequence, that the equations of conservation of momentum and energy, corresponding exactly to equations (49) and (49a), hold good for the components of the total energy. This will be shown in § 17.

§ 17. The Laws of Conservation in the General Case

Equation (52) may readily be transformed so that the second term on the right-hand side vanishes. Contract (52) with respect to the indices μ and σ, and after multiplying the resulting equation by $\frac{1}{2}\delta^\sigma_\mu$, subtract it from equation (52). This gives

$$\frac{\partial}{\partial x_\alpha}(g^{\sigma\beta}\Gamma^\alpha_{\mu\beta} - \tfrac{1}{2}\delta^\sigma_\mu g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}) = -\kappa(t^\sigma_\mu + T^\sigma_\mu). \qquad (52a)$$

On this equation we perform the operation $\partial/\partial x_\sigma$. We have

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left(g^{\sigma}\Gamma^{\alpha}_{\beta\mu}\right) = -\frac{1}{2}\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left[g^{\sigma\beta}g^{\alpha\lambda}\left(\frac{\partial g_{\mu\lambda}}{\partial x_\beta} + \frac{\partial g_{\beta\lambda}}{\partial x_\mu} - \frac{\partial g_{\mu\beta}}{\partial x_\lambda}\right)\right].$$

The first and third terms of the round brackets yield contributions which cancel one another, as may be seen by interchanging, in the contribution of the third term, the summation indices α and σ on the one hand, and β and λ on the other. The second term may be re-modelled by (31), so that we have

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left(g^{\sigma\beta}\Gamma^\alpha_{\mu\beta}\right) = \frac{1}{2}\frac{\partial^3 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta \partial x_\mu} \quad . \quad . \quad (54)$$

The second term on the left-hand side of (52a) yields in the first place

$$-\tfrac{1}{2}\frac{\partial^2}{\partial x_\alpha \partial x_\mu}\left(g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}\right)$$

or

$$\tfrac{1}{4}\frac{\partial^2}{\partial x_\alpha \partial x_\mu}\left[g^{\lambda\beta}g^{\alpha\delta}\left(\frac{\partial g_{\delta\lambda}}{\partial x_\beta}+\frac{\partial g_{\delta\beta}}{\partial x_\lambda}-\frac{\partial g_{\lambda\beta}}{\partial x_\delta}\right)\right].$$

With the choice of co-ordinates which we have made, the term deriving from the last term in round brackets disappears by reason of (29). The other two may be combined, and together, by (31), they give

$$-\tfrac{1}{2}\frac{\partial^3 g^{\alpha\beta}}{\partial x_\alpha \partial x_\beta \partial x_\mu},$$

so that in consideration of (54), we have the identity

$$\frac{\partial^2}{\partial x_\alpha \partial x_\sigma}\left(g^{\rho\beta}\Gamma_{\mu\beta} - \tfrac{1}{2}\delta^\sigma_\mu g^{\lambda\beta}\Gamma^\alpha_{\lambda\beta}\right) \equiv 0 \quad . \quad . \quad (55)$$

From (55) and (52a), it follows that

$$\frac{\partial(t^\sigma_\mu + T^\sigma_\mu)}{\partial x_\sigma} = 0. \quad . \quad . \quad . \quad . \quad (56)$$

Thus it results from our field equations of gravitation that the laws of conservation of momentum and energy are satisfied. This may be seen most easily from the consideration which leads to equation (49a); except that here, instead of the energy components t^σ of the gravitational field, we have to introduce the totality of the energy components of matter and gravitational field.

§ 18. The Laws of Momentum and Energy for Matter, as a Consequence of the Field Equations

Multiplying (53) by $\partial g^{\mu\nu}/\partial x_\sigma$, we obtain, by the method adopted in § 15, in view of the vanishing of

$$g_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma},$$

the equation

[26] $$\frac{\partial t^\alpha_\sigma}{\partial x_\alpha} + \tfrac{1}{2}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}T_{\mu\nu} = 0,$$

or, in view of (56),

$$\frac{\partial T^\alpha_\sigma}{\partial x_\alpha} + \tfrac{1}{2}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}T_{\mu\nu} = 0 \quad . \quad . \quad . \quad (57)$$

Comparison with (41b) shows that with the choice of system of co-ordinates which we have made, this equation predicates nothing more or less than the vanishing of divergence of the material energy-tensor. Physically, the occurrence of the second term on the left-hand side shows that laws of conservation of momentum and energy do not apply in the strict sense for matter alone, or else that they apply only when the $g^{\mu\nu}$ are constant, i.e. when the field intensities of gravitation vanish. This second term is an expression for momentum, and for energy, as transferred per unit of volume and time from the gravitational field to matter. This is brought out still more clearly by re-writing (57) in the sense of (41) as

$$\frac{\partial T^\alpha_\sigma}{\partial x_\alpha} = -\Gamma^\beta_{\alpha\sigma}T^\alpha_\beta \quad . \quad . \quad . \quad (57a)$$

The right side expresses the energetic effect of the gravitational field on matter.

Thus the field equations of gravitation contain four conditions which govern the course of material phenomena. They give the equations of material phenomena completely, if the latter is capable of being characterized by four differential equations independent of one another.*

D. Material Phenomena

The mathematical aids developed in part B enable us forthwith to generalize the physical laws of matter (hydrodynamics, Maxwell's electrodynamics), as they are formulated in the special theory of relativity, so that they will fit in with the general theory of relativity. When this is done, the general principle of relativity does not indeed afford us a further limitation of possibilities; but it makes us acquainted with the influence of the gravitational field on all processes,

* On this question cf. H. Hilbert, Nachr. d. K. Gesellsch. d. Wiss. zu Göttingen, Math.-phys. Klasse, 1915, p. 3. [27]

without our having to introduce any new hypothesis whatever.

Hence it comes about that it is not necessary to introduce definite assumptions as to the physical nature of matter (in the narrower sense). In particular it may remain an open question whether the theory of the electromagnetic field in conjunction with that of the gravitational field furnishes a sufficient basis for the theory of matter or not. The general postulate of relativity is unable on principle to tell us anything about this. It must remain to be seen, during the working out of the theory, whether electromagnetics and the doctrine of gravitation are able in collaboration to perform what the former by itself is unable to do.

§ 19. Euler's Equations for a Frictionless Adiabatic Fluid

Let p and ρ be two scalars, the former of which we call the "pressure," the latter the "density" of a fluid; and let an equation subsist between them. Let the contravariant symmetrical tensor

$$T^{\alpha\beta} = -g^{\alpha\beta}p + \rho\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} \quad . \quad . \quad . \quad (58)$$

be the contravariant energy-tensor of the fluid. To it belongs the covariant tensor

$$T_{\mu\nu} = -g_{\mu\nu}p + g_{\mu\alpha}g_{\mu\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds}\rho, \quad . \quad . \quad (58a)$$

as well as the mixed tensor*

$$T_\sigma^\alpha = -\delta_\sigma^\alpha p + g_{\sigma\beta}\frac{dx_\beta}{ds}\frac{dx_\alpha}{ds}\rho \quad . \quad . \quad (58b)$$

Inserting the right-hand side of (58b) in (57a), we obtain the Eulerian hydrodynamical equations of the general theory of relativity. They give, in theory, a complete solution of the problem of motion, since the four equations (57a), together

* For an observer using a system of reference in the sense of the special theory of relativity for an infinitely small region, and moving with it, the density of energy T_4^4 equals $\rho - p$. This gives the definition of ρ. Thus ρ is not constant for an incompressible fluid.

with the given equation between p and ρ, and the equation

$$g_{\alpha\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 1,$$

are sufficient, $g_{\alpha\beta}$ being given, to define the six unknowns

$$p,\ \rho,\ \frac{dx_1}{ds},\ \frac{dx_2}{ds},\ \frac{dx_3}{ds},\ \frac{dx_4}{ds}.$$

If the $g_{\mu\nu}$ are also unknown, the equations (53) are brought in. These are eleven equations for defining the ten functions $g_{\mu\nu}$, so that these functions appear over-defined. We must remember, however, that the equations (57a) are already contained in the equations (53), so that the latter represent only seven independent equations. There is good reason for this lack of definition, in that the wide freedom of the choice of co-ordinates causes the problem to remain mathematically undefined to such a degree that three of the functions of space may be chosen at will.*

§ 20. Maxwell's Electromagnetic Field Equations for Free Space [28]

Let ϕ_ν be the components of a covariant vector—the electromagnetic potential vector. From them we form, in accordance with (36), the components $F_{\rho\sigma}$ of the covariant six-vector of the electromagnetic field, in accordance with the system of equations

$$F_{\rho\sigma} = \frac{\partial\phi_\rho}{\partial x_\sigma} - \frac{\partial\phi_\sigma}{\partial x_\rho} \quad . \quad . \quad . \quad . \quad (59)$$

It follows from (59) that the system of equations

$$\frac{\partial F_{\rho\sigma}}{\partial x_\tau} + \frac{\partial F_{\sigma\tau}}{\partial x_\rho} + \frac{\partial F_{\tau\rho}}{\partial x_\sigma} = 0 \quad . \quad . \quad . \quad (60)$$ [29]

is satisfied, its left side being, by (37), an antisymmetrical tensor of the third rank. System (60) thus contains essentially four equations which are written out as follows:—

* On the abandonment of the choice of co-ordinates with $g = -1$, there remain *four* functions of space with liberty of choice, corresponding to the four arbitrary functions at our disposal in the choice of co-ordinates.

$$\left.\begin{aligned}\frac{\partial F_{23}}{\partial x_4}+\frac{\partial F_{34}}{\partial x_2}+\frac{\partial F_{42}}{\partial x_3}=0\\ \frac{\partial F_{34}}{\partial x_1}+\frac{\partial F_{41}}{\partial x_3}+\frac{\partial F_{13}}{\partial x_4}=0\\ \frac{\partial F_{41}}{\partial x_2}+\frac{\partial F_{12}}{\partial x_4}+\frac{\partial F_{24}}{\partial x_1}=0\\ \frac{\partial F_{12}}{\partial x_3}+\frac{\partial F_{23}}{\partial x_1}+\frac{\partial F_{31}}{\partial x_2}=0\end{aligned}\right\} \quad . \quad . \quad (60a)$$

This system corresponds to the second of Maxwell's systems of equations. We recognize this at once by setting

$$\left.\begin{aligned}F_{23}=H_x,\ F_{14}=E_x\\ F_{31}=H_y,\ F_{24}=E_y\\ F_{12}=H_z,\ F_{34}=E_z\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (61)$$

Then in place of (60a) we may set, in the usual notation of three-dimensional vector analysis,

$$\left.\begin{aligned}-\frac{\partial H}{\partial t}=\operatorname{curl} E\\ \operatorname{div} H=0\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (60b)$$

We obtain Maxwell's first system by generalizing the form given by Minkowski. We introduce the contravariant six-vector associated with $F^{\alpha\beta}$

[30]

$$F^{\mu\nu}=g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta} \quad . \quad . \quad . \quad . \quad (62)$$

and also the contravariant vector J^{μ} of the density of the electric current. Then, taking (40) into consideration, the following equations will be invariant for any substitution whose invariant is unity (in agreement with the chosen co-ordinates):—

$$\frac{\partial}{\partial x_\nu}F^{\mu\nu}=J^{\mu} \quad . \quad . \quad . \quad . \quad . \quad (63)$$

Let

$$\left.\begin{aligned}F^{23}=H'_x,\ F^{14}=-E'_x\\ F^{31}=H'_y,\ F^{24}=-E'_y\\ F^{12}=H'_z,\ F^{34}=-E'_z\end{aligned}\right\} \quad . \quad . \quad (64)$$

which quantities are equal to the quantities $H_x \ldots E^z$ in the special case of the restricted theory of relativity; and in addition

$$J^1=j_x,\ J^2=j_y,\ J^3=j_z,\ J^4=\rho,$$

we obtain in place of (63)

$$\left.\begin{aligned}\frac{\partial E'}{\partial t}+j=\operatorname{curl} H'\\ \operatorname{div} E'=\rho\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (63a)$$

The equations (60), (62), and (63) thus form the generalization of Maxwell's field equations for free space, with the convention which we have established with respect to the choice of co-ordinates.

The Energy-components of the Electromagnetic Field.— We form the inner product

$$\kappa_\sigma=F_{\sigma\mu}J^{\mu} \quad . \quad . \quad . \quad . \quad . \quad (65)$$

By (61) its components, written in the three-dimensional manner, are

$$\left.\begin{aligned}\kappa_1&=\rho E_x+[j\,.\,H]^x\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\\ &\cdot\quad\cdot\quad\cdot\quad\cdot\\ \kappa_4&=-(jE)\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (65a)$$

κ_σ is a covariant vector the components of which are equal to the negative momentum, or, respectively, the energy, which is transferred from the electric masses to the electromagnetic field per unit of time and volume. If the electric masses are free, that is, under the sole influence of the electromagnetic field, the covariant vector κ_σ will vanish.

To obtain the energy-components T_σ^ν of the electromagnetic field, we need only give to equation $\kappa_\sigma=0$ the form of equation (57). From (63) and (65) we have in the first place

$$\kappa_\sigma=F_{\sigma\mu}\frac{\partial F^{\mu\nu}}{\partial x_\nu}=\frac{\partial}{\partial x_\nu}(F_{\sigma\mu}F^{\mu\nu})-F^{\mu\nu}\frac{\partial F_{\sigma\mu}}{\partial x_\nu}.$$

The second term of the right-hand side, by reason of (60), permits the transformation

$$F^{\mu\nu}\frac{\partial F_{\sigma\mu}}{\partial x_\nu}=-\tfrac{1}{2}F^{\mu\nu}\frac{\partial F_{\mu\nu}}{\partial x_\sigma}=-\tfrac{1}{2}g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}\frac{\partial F_{\mu\nu}}{\partial x_\sigma},$$

which latter expression may, for reasons of symmetry, also be written in the form

$$-\tfrac{1}{4}\left[g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}\frac{\partial F_{\mu\nu}}{\partial x_\sigma} + g^{\mu\alpha}g^{\nu\beta}\frac{\partial F_{\alpha\beta}}{\partial x_\sigma}F_{\mu\nu}\right].$$

But for this we may set

$$-\tfrac{1}{4}\frac{\partial}{\partial x_\sigma}(g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}F_{\mu\nu}) + \tfrac{1}{4}F_{\alpha\beta}F_{\mu\nu}\frac{\partial}{\partial x_\sigma}(g^{\mu\alpha}g^{\nu\beta}).$$

The first of these terms is written more briefly

$$-\tfrac{1}{4}\frac{\partial}{\partial x_\sigma}(F^{\mu\nu}F_{\mu\nu});$$

the second, after the differentiation is carried out, and after some reduction, results in

[31] $$-\tfrac{1}{2}F^{\mu\tau}F_{\mu\nu}g^{\nu\rho}\frac{\partial g_{\sigma\tau}}{\partial x_\sigma}.$$

Taking all three terms together we obtain the relation

$$\kappa_\sigma = \frac{\partial T_\sigma^\nu}{\partial x_\nu} - \tfrac{1}{2}g^{\tau\mu}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}T_\tau^\nu \quad . \quad . \quad . \quad (66)$$

where

$$T_\sigma^\nu = -F_{\sigma\alpha}F^{\nu\alpha} + \tfrac{1}{4}\delta_\sigma^\nu F_{\alpha\beta}F^{\alpha\beta}.$$

Equation (66), if κ_σ vanishes, is, on account of (30), equivalent to (57) or (57*a*) respectively. Therefore the T_σ^ν are the energy-components of the electromagnetic field. With the help of (61) and (64), it is easy to show that these energy-components of the electromagnetic field in the case of the special theory of relativity give the well-known Maxwell-Poynting expressions.

We have now deduced the general laws which are satisfied by the gravitational field and matter, by consistently using a system of co-ordinates for which $\sqrt{-g} = 1$. We have thereby achieved a considerable simplification of formulæ and calculations, without failing to comply with the requirement of general covariance; for we have drawn our equations from generally covariant equations by specializing the system of co-ordinates.

Still the question is not without a formal interest, whether with a correspondingly generalized definition of the energy-components of gravitational field and matter, even without specializing the system of co-ordinates, it is possible to formulate laws of conservation in the form of equation (56), and field equations of gravitation of the same nature as (52) or (52a), in such a manner that on the left we have a divergence (in the ordinary sense), and on the right the sum of the energy-components of matter and gravitation. I have found that in both cases this is actually so. But I do not think that the communication of my somewhat extensive reflexions on this subject would be worth while, because after all they do not give us anything that is materially new.

E

§ 21. Newton's Theory as a First Approximation [32]

As has already been mentioned more than once, the special theory of relativity as a special case of the general theory is characterized by the $g_{\mu\nu}$ having the constant values (4). From what has already been said, this means complete neglect of the effects of gravitation. We arrive at a closer approximation to reality by considering the case where the $g_{\mu\nu}$ differ from the values of (4) by quantities which are small compared with 1, and neglecting small quantities of second and higher order. (First point of view of approximation.)

It is further to be assumed that in the space-time territory under consideration the $g_{\mu\nu}$ at spatial infinity, with a suitable choice of co-ordinates, tend toward the values (4); i.e. we are considering gravitational fields which may be regarded as generated exclusively by matter in the finite region.

It might be thought that these approximations must lead us to Newton's theory. But to that end we still need to approximate the fundamental equations from a second point of view. We give our attention to the motion of a material point in accordance with the equations (16). In the case of the special theory of relativity the components

$$\frac{dx_1}{ds}, \frac{dx_2}{ds}, \frac{dx_3}{ds}$$

may take on any values. This signifies that any velocity

$$v = \sqrt{\left(\frac{dx_1}{dx_4}\right)^2 + \left(\frac{dx_2}{dx_4}\right)^2 + \left(\frac{dx_3}{dx_4}\right)^2}$$

may occur, which is less than the velocity of light *in vacuo*. If we restrict ourselves to the case which almost exclusively offers itself to our experience, of v being small as compared with the velocity of light, this denotes that the components

$$\frac{dx_1}{ds}, \frac{dx_2}{ds}, \frac{dx_3}{ds}$$

are to be treated as small quantities, while dx_4/ds, to the second order of small quantities, is equal to one. (Second point of view of approximation.)

Now we remark that from the first point of view of approximation the magnitudes $\Gamma^{\tau}_{\mu\nu}$ are all small magnitudes of at least the first order. A glance at (46) thus shows that in this equation, from the second point of view of approximation, we have to consider only terms for which $\mu = \nu = 4$. Restricting ourselves to terms of lowest order we first obtain in place of (46) the equations

$$\frac{d^2x_\tau}{dt^2} = \Gamma^{\tau}_{44}$$

where we have set $ds = dx_4 = dt$; or with restriction to terms which from the first point of view of approximation are of first order:—

$$\frac{d^2x_\tau}{dt^2} = [44, \tau] \quad (\tau = 1, 2, 3)$$

$$\frac{d^2x_4}{dt^2} = -[44, 4].$$

If in addition we suppose the gravitational field to be a quasi-static field, by confining ourselves to the case where the motion of the matter generating the gravitational field is but slow (in comparison with the velocity of the propagation of light), we may neglect on the right-hand side differentiations with respect to the time in comparison with those with respect to the space co-ordinates, so that we have

$$\frac{d^2x_\tau}{dt^2} = -\tfrac{1}{2}\frac{\partial g_{44}}{\partial x_\tau} \quad (\tau = 1, 2, 3) \quad . \quad . \quad (67)$$

This is the equation of motion of the material point according to Newton's theory, in which $\tfrac{1}{2}g_{44}$ plays the part of the gravitational potential. What is remarkable in this result is that the component g_{44} of the fundamental tensor alone defines, to a first approximation, the motion of the material point.

We now turn to the field equations (53). Here we have to take into consideration that the energy-tensor of "matter" is almost exclusively defined by the density of matter in the narrower sense, i.e. by the second term of the right-hand side of (58) [or, respectively, (58a) or (58b)]. If we form the approximation in question, all the components vanish with the one exception of $T_{44} = \rho = T$. On the left-hand side of (53) the second term is a small quantity of second order; the first yields, to the approximation in question,

$$\frac{\partial}{\partial x_1}[\mu\nu, 1] + \frac{\partial}{\partial x_2}[\mu\nu, 2] + \frac{\partial}{\partial x_3}[\mu\nu, 3] - \frac{\partial}{\partial x_4}[\mu\nu, 4].$$

For $\mu = \nu = 4$, this gives, with the omission of terms differentiated with respect to time,

$$-\tfrac{1}{2}\left(\frac{\partial^2 g_{44}}{\partial x_1^2} + \frac{\partial^2 g_{44}}{\partial x_2^2} + \frac{\partial^2 g_{44}}{\partial x_3^2}\right) = -\tfrac{1}{2}\nabla^2 g_{44}.$$

The last of equations (53) thus yields

$$\nabla^2 g_{44} = \kappa\rho \quad . \quad . \quad . \quad . \quad (68)$$

The equations (67) and (68) together are equivalent to Newton's law of gravitation.

By (67) and (68) the expression for the gravitational potential becomes

$$-\frac{\kappa}{8\pi}\int\frac{\rho d\tau}{r} \quad . \quad . \quad . \quad . \quad (68a)$$

while Newton's theory, with the unit of time which we have chosen, gives

$$-\frac{K}{c^2}\int\frac{\rho d\tau}{r}$$

in which K denotes the constant $6{\cdot}7 \times 10^{-8}$, usually called the constant of gravitation. By comparison we obtain

$$\kappa = \frac{8\pi K}{c^2} = 1{\cdot}87 \times 10^{-27} \quad . \quad . \quad (69)$$

§ 22. Behaviour of Rods and Clocks in the Static Gravitational Field. Bending of Light-rays. Motion of the Perihelion of a Planetary Orbit

To arrive at Newton's theory as a first approximation we had to calculate only one component, g_{44}, of the ten $g_{\mu\nu}$ of the gravitational field, since this component alone enters into the first approximation, (67), of the equation for the motion of the material point in the gravitational field. From this, however, it is already apparent that other components of the $g_{\mu\nu}$ must differ from the values given in (4) by small quantities of the first order. This is required by the condition $g = -1$.

For a field-producing point mass at the origin of co-ordinates, we obtain, to the first approximation, the radially symmetrical solution

$$\left.\begin{array}{ll} g_{\rho\sigma} = -\delta_{\rho\sigma} - \alpha\dfrac{x_\rho x_\sigma}{r^3} & (\rho, \sigma = 1, 2, 3) \\ g_{\rho 4} = g_{4\rho} = 0 & (\rho = 1, 2, 3) \\ g_{44} = 1 - \dfrac{\alpha}{r} & \end{array}\right\} \quad . \quad (70)$$

[33] where $\delta_{\rho\sigma}$ is 1 or 0, respectively, accordingly as $\rho = \sigma$ or $\rho \neq \sigma$, and r is the quantity $+\sqrt{x_1^2 + x_2^2 + x_3^2}$. On account of (68a)

$$\alpha = \frac{\kappa M}{4\pi}, \quad . \quad . \quad . \quad . \quad (70a)$$

if M denotes the field-producing mass. It is easy to verify that the field equations (outside the mass) are satisfied to the first order of small quantities.

We now examine the influence exerted by the field of the mass M upon the metrical properties of space. The relation

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu.$$

always holds between the "locally" (§ 4) measured lengths and times ds on the one hand, and the differences of co-ordinates dx_ν on the other hand.

For a unit-measure of length laid "parallel" to the axis of x, for example, we should have to set $ds^2 = -1$; $dx_2 = dx_3 = dx_4 = 0$. Therefore $-1 = g_{11}dx_1^2$. If, in addition, the unit-measure lies on the axis of x, the first of equations (70) gives

$$g_{11} = -\left(1 + \frac{\alpha}{r}\right).$$

From these two relations it follows that, correct to a first order of small quantities,

$$dx = 1 - \frac{\alpha}{2r} \quad . \quad . \quad . \quad . \quad (71)$$

The unit measuring-rod thus appears a little shortened in relation to the system of co-ordinates by the presence of the gravitational field, if the rod is laid along a radius.

In an analogous manner we obtain the length of co-ordinates in tangential direction if, for example, we set
$ds^2 = -1$; $dx_1 = dx_3 = dx_4 = 0$; $x_1 = r$, $x_2 = x_3 = 0$.
The result is

$$-1 = g_{22}dx_2^2 = -dx_2^2 \quad . \quad . \quad . \quad (71a)$$

With the tangential position, therefore, the gravitational field of the point of mass has no influence on the length of a rod.

Thus Euclidean geometry does not hold even to a first approximation in the gravitational field, if we wish to take one and the same rod, independently of its place and orientation, as a realization of the same interval; although, to be sure, a glance at (70a) and (69) shows that the deviations to be expected are much too slight to be noticeable in measurements of the earth's surface.

Further, let us examine the rate of a unit clock, which is arranged to be at rest in a static gravitational field. Here we have for a clock period $ds = 1$; $dx_1 = dx_2 = dx_3 = 0$
Therefore

$$1 = g_{44}dx_4^2;$$

$$dx_4 = \frac{1}{\sqrt{g_{44}}} = \frac{1}{\sqrt{(1 + (g_{44} - 1))}} = 1 - \tfrac{1}{2}(g_{44} - 1)$$

or

$$dx_4 = 1 + \frac{\kappa}{8\pi}\int \rho\frac{d\tau}{r} \quad . \quad . \quad . \quad . \quad (72)$$

Thus the clock goes more slowly if set up in the neighbourhood of ponderable masses. From this it follows that the spectral lines of light reaching us from the surface of large stars must appear displaced towards the red end of the spectrum.*

We now examine the course of light-rays in the static gravitational field. By the special theory of relativity the velocity of light is given by the equation

$$-dx_1^2 - dx_2 - dx_3^2 + dx_4^2 = 0$$

and therefore by the general theory of relativity by the equation

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu = 0 \quad . \quad . \quad . \quad . \quad (73)$$

If the direction, i.e. the ratio $dx_1 : dx_2 : dx_3$ is given, equation (73) gives the quantities

$$\frac{dx_1}{dx_4}, \frac{dx_2}{dx_4}, \frac{dx_3}{dx_4}$$

and accordingly the velocity

$$\sqrt{\left(\frac{dx_1}{dx_4}\right)^2 + \left(\frac{dx_2}{dx_4}\right)^2 + \left(\frac{dx_3}{dx_4}\right)^2} = \gamma$$

defined in the sense of Euclidean geometry. We easily recognize that the course of the light-rays must be bent with regard to the system of co-ordinates, if the $g_{\mu\nu}$ are not constant. If n is a direction perpendicular to the propagation of light, the Huyghens principle shows that the light-ray, en-
[35] visaged in the plane (γ, n), has the curvature $-\partial\gamma/\partial n$.

We examine the curvature undergone by a ray of light passing by a mass M at the distance Δ. If we choose the system of co-ordinates in agreement with the accompanying diagram, the total bending of the ray (calculated positively if

* According to E. Freundlich, spectroscopical observations on fixed stars of
[34] certain types indicate the existence of an effect of this kind, but a crucial test of this consequence has not yet been made.

concave towards the origin) is given in sufficient approximation by

$$B = \int_{-\infty}^{+\infty} \frac{\partial\gamma}{\partial x_1} dx_2,$$

while (73) and (70) give

$$\gamma = \sqrt{\left(-\frac{g_{44}}{g_{22}}\right)} = 1 - \frac{a}{2r}\left(1 + \frac{x_2^2}{r^2}\right).$$

Carrying out the calculation, this gives

$$B = \frac{2a}{\Delta} = \frac{\kappa M}{2\pi\Delta}. \quad . \quad . \quad . \quad . \quad (74)$$ [36]

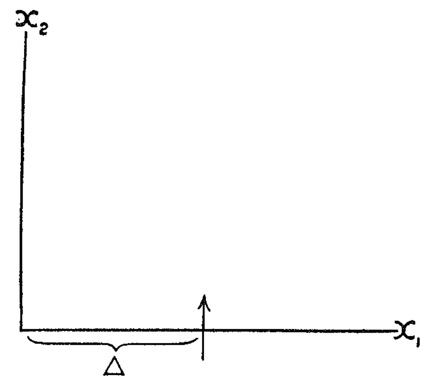


FIG. 8.

According to this, a ray of light going past the sun undergoes a deflexion of 1·7″; and a ray going past the planet Jupiter a deflexion of about ·02″.

If we calculate the gravitational field to a higher degree of approximation, and likewise with corresponding accuracy the orbital motion of a material point of relatively infinitely small mass, we find a deviation of the following kind from the Kepler-Newton laws of planetary motion. The orbital ellipse of a planet undergoes a slow rotation, in the direction of motion, of amount

$$\epsilon = 24\pi^3\frac{a^2}{T^2c^2(1-e^2)} \quad . \quad . \quad . \quad . \quad (75)$$

per revolution. In this formula a denotes the major semi-axis, c the velocity of light in the usual measurement, e the eccentricity, T the time of revolution in seconds.*

Calculation gives for the planet Mercury a rotation of the orbit of 43″ per century, corresponding exactly to astronomical observation (Leverrier); for the astronomers have discovered in the motion of the perihelion of this planet, after allowing for disturbances by other planets, an inexplicable remainder of this magnitude.

[37] * For the calculation I refer to the original papers: A. Einstein, Sitzungsber. d. Preuss. Akad. d. Wiss., 1915, p. 831; K. Schwarzschild, *ibid.*, 1916, p. 189.

Doc. 31

Appendix. Formulation of the Theory on the Basis of a Variational Principle

Not translated for this volume.

Doc. 32

Session of the physical-mathematical class on June 22, 1916 [p. 688]

Approximative Integration of the Field Equations of Gravitation

by A. Einstein

For the treatment of the special (not basic) problems in gravitational theory one can be satisfied with a first approximation of the $g_{\mu\nu}$. The same reasons as in the special theory of relativity make it advantageous to use the imaginary time variable $x_4 = it$. By "first approximation" we mean that the quantities $\gamma_{\mu\nu}$, defined by the equation

$$g_{\mu\nu} = -\delta_{\mu\nu} + \gamma_{\mu\nu}, \tag{1}$$

are small compared to 1, such that their squares and products are negligible compared with first powers; furthermore, they have a tensorial character under linear, orthogonal transformations. In addition, $\delta_{\mu\nu} = 1$ or $\delta_{\mu\nu} = 0$ resp. depending upon $\mu = \nu$ or $\mu \neq \nu$.

We shall show that these $\gamma_{\mu\nu}$ can be calculated in a manner analogous to that of retarded potentials in electrodynamics. From this follows next that gravitational fields propagate at the speed of light. Subsequent to this general solution we shall investigate gravitational waves and how they originate. It turned out that my suggested choice of a system of reference with the condition $g = |g_{\mu\nu}| = -1$ is not advantageous for the calculation of fields in first approximation. A note in a letter [1] from the astronomer DE SITTER alerted me to his finding that a choice of reference system, different from the one I had previously given,[1] leads to a simpler expression of the gravitational field of a mass point at rest. I therefore take the generally invariant field equations as a basis in what follows.

§1. Integration of the Approximated Equations of the Gravitational Field [p. 689]

The field equations in their covariant form are [3]

[1] *Sitzungsber.* 47 (1915), p. 833. [2]

The Quantum Theory of Radiation

A. Einstein
(Received March, 1917)

The formal similarity of the spectral distribution curve of temperature radiation to Maxwell's velocity distribution curve is too striking to have remained hidden very long. Indeed, in the important theoretical paper in which Wien derived his displacement law

$$\rho = \nu^3 f\left(\frac{\nu}{T}\right) \tag{1}$$

he was led by this similarity to a farther correspondence with the radiation formula. He discovered, as is known, the formula [Wien's radiation formula]

$$\rho = \alpha\nu^3 e^{-\frac{h\nu}{kT}} \tag{2}$$

which is recognized today as the correct limiting formula for large values of $\frac{\nu}{T}$. Today we known that no consideration which is based on classical mechanics and electrodynamics can lead to a useful radiation formula; rather that the classical theory leads to the Rayleigh formula.

$$\rho = \frac{k\alpha}{h}\,\nu^2 T \tag{3}$$

After Planck, in his ground–breaking investigation, established his radiation formula

$$\rho = \alpha\nu^3\,\frac{1}{e^{\frac{h\nu}{kT}} - 1} \tag{4}$$

on the assumption that there are discrete elements of energy, from which quantum theory developed very rapidly, Wien's considerations, from which formula (2) evolved, quite naturally were forgotten.

A little while ago I obtained a derivation, related to Wien's original idea, of the Planck radiation formula which is based on the fundamental assumption of quantum theory and which makes use of the relationship of Maxwell's curve to the spectral distribution curve. This derivation deserves consideration not only because of its simplicity, but especially because it appears to clarify the processes of emission and absorption of radiation in matter, which is still in such darkness for us. In setting down certain fundamental hypotheses concerning the absorption and emission of radiation by molecules that are closely related to quantum theory, I showed that molecules with a distribution of states in the quantum theoretical sense for temperature equilibrium are in dynamical equilibrium with the Planck radiation; in this way, the Planck formula (4) was obtained in a surprisingly simple and general way. It was obtained from the condition that the quantum theoretic partition of states of the internal energy of the molecules is established only by the emission and absorption of radiation.

If the assumed hypotheses about the interaction of matter and radiation are correct, they will give us more than just the correct statistical partition or distribution of the internal energy of the molecules. During absorption and emission of radiation there is also present a transfer of momentum to the molecules; this means that just the interaction of radiation and molecules leads to a velocity distribution of the latter. This must early be the same as the velocity distribution which molecules acquire as the result of their mutual interaction by collisions, that is, it must coincide with the Maxwell distribution. we must require that the mean kinetic energy which a molecule (per degree of freedom) acquires in a Plank radiation field of temperature T be

$$\frac{kT}{2};$$

this must be valid regardless of the nature of the molecules and independent of frequencies which the molecules absorb and emit. In this paper we wish to verify that this far–reaching requirement is, indeed, satisfied quite generally; as a result of this our simple hypotheses about the emission and absorption of radiation acquire new supports.

In order to obtain this result, however, we must enlarge, in a definite way, the previous fundamental hypothesis which were related entirely to the exchange of energy. We are faced with this question; Does the molecule suffer a push, when it absorbs or emits the energy ϵ? As an example we consider, from the classical point of view, the emission of radiation. If a body emits the energy ϵ, it acquires a backward thrust [impulse] $\frac{\epsilon}{c}$ if all the radiation ϵ is radiated in the same direction. If, however, the radiation occurs through a spatially symmetric process, for example, spherical waves. there is then no recoil at all. This alternative also plays a role in the quantum theory of radiation. If a molecule, in going from one possible quantum theoretic state to another, absorbs or emits the energy ϵ in the form of radiation, such an elementary process can be looked upon as partly or fully directed in space, or also as a symmetric (non–directed) one. It turns out that we obtain a theory that is free of contradictions only if we consider the above elementary processes as being fully directed events; herein lies the principal result of the considerations that follow.

Fundamental Hypotheses of the Quantum Theory–Canonical Distribution of State

According to the quantum theory, a molecule of a definite kind may, aside from its orientation and its translational motion, be in one only a discrete set of states $Z_1, Z_2, \ldots Z_n \ldots$ whose (internal) energies are $\epsilon_1, \epsilon_2, \ldots \epsilon_n \ldots$. If the molecules of this kind belong to a gas of temperature T, then the relative abundance W_n of the state Z_n is given by the statistical mechanical canonical partition function for states

$$W_n = p_n e^{-\frac{\epsilon_n}{kT}} \qquad (5)$$

In this formula $k = \frac{R}{N}$ is the well-known Boltzmann constant, p_n a number that is independent of T and characteristic of the molecule and the state, which we may call the statistical "weight" of the state. Formula (5) can be derived from the Boltzmann principle or purely from thermodynamics. Equation (5) is the expression of the most far–reaching generalization of the Maxwellian distribution of velocities.

The latest important advances in quantum theory deal with the theoretical determination of the quantum theoretical possible states Z_n and their weights p_n. For the principal part of the present investigation, it is not necessary to have a more detailed determination of the quantum states.

Hypotheses about the Energy Exchange Through Radiation

Let Z_n and Z_m be two possible quantum theoretical states of a gas molecule whose energies ϵ_n and ϵ_m respectively, satisfy the inequality

$$\epsilon_m > \epsilon_n$$

Let the molecule be able to pass from the state Z_n to the state Z_m by absorbing the radiation energy $\epsilon_m - \epsilon_n$, similarly let the transition from state Z_n to the state Z_m be possible through the emission of this amount of energy. Let the radiation emitted or absorbed by the molecule for the given index and combination (m, n) have the characteristic frequency ν.

We now introduce certain hypotheses about the laws which are decisive for these transitions. These hypotheses are obtained by carrying over the known classical relations for a Planck resonator to the unknown quantum theoretical relations.

Emission

A Planck resonator that is vibrating radiates energy according to Hertz, in a known way independently of whether it is stimulated by an external field or not. In accordance with this, let a molecule be able to pass from the state Z_m to the state Z_n with the emission of radiant energy $\epsilon_m - \epsilon_n$ of frequency ν without being excited by any external cause. Let the probability dW for this to happen in the time dt be

$$dW = A_m^n \; dt \tag{A}$$

where A_m^n is a characteristic constant for the given index combination.

The assumed statistical law corresponds to that of a radioactive reaction: that elementary process of such a reaction in which only γ-rays are emitted. We need not assume that this process requires no time; this time need only be negligible compared to the times which the molecule spends in the states Z_1, and so on.

Incident Radiation

If a Planck resonator is in a radiation field, the energy of the resonator changes because the electromagnetic field of the radiation does work on the resonator; this work can be positive or negative depending on the phases of the resonator and the oscillating field. In accordance with this, we introduce the following quantum theoretical hypothesis. Under the action of the radiation density ρ of the frequency ν a molecule in state Z_n can go over to state Z_m absorbing the radiation energy $\epsilon_m - \epsilon_n$ in accordance with the probability law

$$dW = B_n^m \ \rho \ dt \tag{B}$$

In the same way, let the transition $Z_m \to Z_n$ under the action of the radiation also be possible, whereby the radiation energy $\epsilon_m - \epsilon_n$ is emitted according to the probability law

$$dW = B_m^n \ \rho \ dt \tag{B'}$$

B_n^m and B_m^n are constants. We call both processes "changes of states through incident radiation."

The question presents itself now as to the momentum that is transferred to the molecule in these changes of state. We begin with the events associated with incident radiation. If a directed bundle of rays does work on a Planck resonator, then an equivalent amount of energy is removed from the bundle. this transfer of energy results, according to the law of momentum, to a momentum transfer from the beam to the resonator. The latter therefore experiences a force in the direction of the ray of the radiation beam. If the energy transferred is negative, the force acting on the resonator is opposite in direction. In the case of the quantum hypothesis, this clearly means the following. If, as the result of incident radiation, the process $Z_n \to Z_m$ occurs, then an amount of momentum

$$\frac{\epsilon_m - \epsilon_n}{c}$$

is transferred to the molecule in the direction of propagation of the bundle of radiation. If we have the process $Z_m \to Z_n$ for the case of incident radiation, the magnitude of the transferred momentum is the same, but it is in the opposite direction. If a molecule is simultaneously exposed to many bundles of radiation, we assume that the total energy $Z_m \to Z_n$ is taken from or added to just one of these bundles, so that even in this case the momentum

$$\frac{\epsilon_m - \epsilon_n}{c}$$

is transferred to the molecule.

In the case of emission of energy by radiation by a Planck resonator, there is no net transfer of momentum to the resonator because, according to classical theory, of emission occurs as a spherical wave. However, we have already noted that we can arrive at a contradiction–free quantum theory only if we assume that the process of emission is a directed one. Every elementary process of emission ($Z_m \to Z_n$) will then result in a transfer to the molecule of an amount of momentum

$$\frac{\epsilon_m - \epsilon_n}{c}.$$

If the molecule is isotropic, we must take every direction of emission as equally probable. If the molecule is not isotropic, we arrive at the same result if the orientation changes in a random way in the course of time. We must, in any case, make such an assumption also for the statistical laws (B) and (B') for incident radiation since otherwise the constants B_n^m and B_m^n would have to depend on direction, which we can avoid by assuming isotropy or pseudo–isotropy (through setting up temporal mean values).

Derivation of the Planck Radiation Law

We now enquire about those effective radiation densities ρ which must prevail in order that the energy exchange between molecules and radiation as a result of the statistical laws (A), (B) and (B') shall not disturb the distribution of molecular states present as a consequence of equation (5). For this, it is necessary and sufficient that on the average, per unit time, as many elementary processes of type (B) take place as processes (A) and (B') together. This condition gives as a result of (5), (A), (B), (B'), for the elementary processes corresponding to the index combination (m, n) the equation

$$p_n e^{-\frac{\epsilon_n}{kT}} B_n^m \rho = p_m e^{-\frac{\epsilon_m}{kT}} (B_m^n \, \rho + A_m^n)$$

If, further, ρ is to become infinite as T does, the constants B_n^m and B_m^n must satisfy the relation

$$p_n B_n^m = p_m B_m^n \tag{6}$$

We then obtain as the condition for dynamical equilibrium the equation

$$\rho = \frac{A_m^n / B_m^n}{e^{\frac{\epsilon_m - \epsilon_n}{kT}} - 1} \tag{7}$$

This is the dependence of the radiation density on the temperature that is given by the Planck law. From the Wien displacement law (1) it then following immediately that

$$\frac{A_m^n}{B_m^n} = \alpha \nu^3 \tag{8}$$

and

$$\epsilon_m - \epsilon_n = h\nu \tag{9}$$

where α and h are universal constants. To obtain the numerical values of α and h we must have an exact theory of electrodynamic and mechanical processes; we content ourselves for the moment with the Rayleigh law in the limit of high temperatures, where the classical theory is valid in the limit.

Equation (9) is, as we know, the second principal rule in Bohr's theory of spectra, about which we may assert, following upon Sommerfeld's and Epstein's completion of the theory, that it belong to the most fully verified domain of our science. It also contain implicitly the photochemical equivalent law, as I have already shown.

Method for Calculating the Motion of Molecules in Radiation Fields

We now turn our attention to the investigation of the motion imparted to our molecules by the radiation field. we make use in this of a method that is known to us from the theory of Brownian motion and which I have often used in investigating motions in a region containing radiation. To simplify the calculation, we shall carry it through for the case in which the motion occurs only along the X-direction of the coordinate system. We further content ourselves with calculating the mean value of the kinetic energy of the translation motion, and thus dispense with proof that these velocities v are distributed according to the Maxwell law. Let the mass M of the molecule be large enough so that higher powers of $\frac{v}{c}$ can be neglected relative to lower ones; we can then apply the usual mechanics to the molecule. Moreover, without any loss in generality, we may carry out the calculation as through the states with indices m and n were the ones the molecule can be in.

The momentum Mv of a molecule undergoes two kinds of changes in the short time τ. Even though the radiation is the same in all directions, the molecule, because of its motion, will experience a resistance to its motion that stems from the radiation. Let this opposing force be Rv, where R is a constant to be determined later. This force would ultimately bring the molecule to rest if the randomness of the action of the radiation field were not such as to transfer to the molecule a momentum Δ of alternating sign and varying magnitude; this random effect will, in opposition to the previous one, sustain a certain amount of motion of the molecule. At the end of the given short time τ the momentum of the molecule will equal

$$Mv - Rv\tau + \Delta$$

Since the velocity distribution is to remain constant in time, the mean of the absolute value of the above quantity must equal that of the quantity Mv; thus, the mean values of the squares of both quantities averaged over a long time or a large number of molecules must be equal:

$$\overline{(Mv - Rv\tau + \Delta)^2} = \overline{(Mv)^2}$$

Since we have taken into account the influence of v on the momentum of the molecule separately, we must discard the mean value $v\Delta$. On developing the left–hand side of the equation we thus obtain

$$\overline{\Delta^2} = \overline{2RMv^2\tau} \tag{10}$$

The mean value $\overline{v^2}$ which the radiation of temperature T by its interaction imparts to the molecule must just equal the mean value $\overline{v^2}$ which the gas molecule acquires at temperature T according to the gas law and the kinetic theory of gases. For otherwise the presence of our molecules would disturb the thermal equilibrium between thermal radiation and an arbitrary gas of the same temperature. We must therefore have

$$\frac{\overline{Mv^2}}{2} = \frac{kT}{2} \tag{11}$$

Equation (10) thus goes over into

$$\frac{\overline{\Delta^2}}{\tau} = 2RkT \tag{12}$$

The investigation is now to be carried through as follows. For a given radiation density ($\rho(\nu)$) we shall be able to compute $\overline{\Delta^2}$ and R by means of our hypotheses about the interaction between radiation and molecules. If we put this result into (12), this equation will have to be identically satisfied when ρ is expressed as a function of ν and T by means of Planck's equation (4).

Computing R

Let a molecule of given kind be in uniform motion with speed v along the X-axis of the coordinate system K. We inquire about the momentum transferred on the average from the radiation to the molecule per unit time. To calculate this we must consider the radiation from a coordinate system K' that is at rest with respect to the given molecule. For we have formulated our hypotheses about emission and absorption only for molecules at rest. The transformation to the system K' has often been performed in the literature. Nevertheless, I shall repeat the simple considerations here for the sake of clarity.

Relative to K the radiation is isotropic, that is, the quantity of radiation in a solid angle $d\kappa$ in the direction of the radiation in a frequency range $d\nu$ is

$$\rho d\nu \ \frac{d\kappa}{4\pi} \tag{13}$$

where ρ depends only on the frequency ν but not on the direction of the radiation. This special beam corresponds to a special beam in the system K' which is also characterized by a frequency range $d\nu'$ and a solid angle $d\kappa'$. The volume density of this special beam is

$$\rho'(\nu', \phi')d\nu' \ \frac{d\kappa'}{4\pi} \tag{13$'$}$$

This defines ρ'. It depends on the direction of the radiation which, in the familiar manner, is defined by the angle ϕ' it makes with the X' axis and which its projection on the Y', Z' plane makes with the Y' axis. These angles correspond to the angles ϕ and ψ which in an analogous manner determine the direction of $d\kappa$ in K.

To begin with, it is clear that the same transformation law between (13) and (13$'$) must hold as between the amplitudes A^2 and A'^2 of a plane wave moving in the corresponding direction. Hence, to our desired approximation we have

$$\frac{\rho'(\nu', \phi')d\nu' d\kappa'}{\rho(\nu)d\nu d\kappa} = 1 - 2 \ \frac{v}{c} \ \cos\phi \tag{14}$$

or

$$\rho'(\nu', \phi') = \rho(\nu)\frac{d\nu}{d\nu'} \ \frac{d\kappa}{d\kappa'} \left(1 - 2 \ \frac{v}{c} \ \cos\phi\right) \tag{14$'$}$$

The relativity theory further gives the formulae, valid to the desired approximation,

$$\nu' = \nu \left(1 - \frac{v}{c} \ \cos\phi\right) \tag{15}$$

$$\cos\phi' = \cos\phi - \frac{v}{c} + \frac{v}{c} \ \cos^2\phi \tag{16}$$

$$\psi' = \psi \tag{17}$$

from (15) it follows, to the same approximation that

$$\nu = \nu' \left(1 + \frac{v}{c}\ \cos\phi'\right)$$

Hence, again to the desired approximation

$$\rho(\nu) = \rho\left(\nu' + \frac{v}{c}\ \nu' \cos\phi'\right)$$

or

$$\rho(\nu) = \rho(\nu') + \frac{\partial\rho(\nu')}{\partial\nu}\left(\frac{v}{c}\ \nu' \cos\phi'\right) \tag{18}$$

Further, according to (15), (16), and (17)

$$\frac{d\nu}{d\nu'} = \left(1 + \frac{v}{c}\ \cos\phi'\right)$$

$$\frac{d\kappa}{d\kappa'} = \frac{\sin\phi\ d\phi\ d\psi}{\sin\phi'\ d\phi'\ d\psi'} = \frac{d(\cos\phi)}{d(\cos\phi')} = 1 - 2\ \frac{v}{c}\ \cos\phi'$$

As a result of these two equations and equation (18), equation (14′) goes over into

$$\rho'(\nu', \phi') = \left[(\rho)_{\nu'} + \frac{v}{c}\ \nu' \cos\phi' \left(\frac{\partial\rho}{\partial\nu}\right)_{\nu'}\right]\left(1 - 3\ \frac{v}{c}\ \cos\phi'\right) \tag{19}$$

With the aid of (19) and our hypotheses about the radiation from and radiation onto molecules, we can easily calculate the average momentum transferred to the molecule per unit time. Before we can do this, however, we must say something to justify our procedure. It may be objected that equations (14), (15), (16) are based on maxwell's theory of the electromagnetic field that is not consistent with the quantum theory. This objection deals, however, more with the form than with the substance of the problem. For, no matter how the theory of electromagnetic processes may be formulated, in any case the Doppler principle and the law of aberration still remain, and hence also the equations (15) and (16). Moreover, the validity of the energy relationship (14) certainly extends beyond that of the wave theory; this transformation law is also valid, for example, according to relativity theory, for the energy density of a mass of infinitesimally small rest density that is moving with the [quasi-] speed of light. We may therefore assert the validity of equation (19) for any theory of radiation.

The radiation belonging to the solid angle $d\kappa'$ would, according to (B), give rise to

$$B_n^m \rho'(\nu', \phi')\ \frac{d\kappa'}{4\pi}$$

elementary processes per second of radiation events of the type $Z_n \to Z_m$ if the molecule after each such process immediately returned to state Z_n. Actually, however, the time of lingering in state Z_n, according to (5), is

$$\frac{1}{S} \cdot p_n e^{-\frac{\epsilon_n}{kT}}$$

where we have used the abbreviation

$$S = p_n e^{-\frac{\epsilon_n}{kT}} + p_m e^{-\frac{\epsilon_m}{kT}} \tag{20}$$

The number of these processes per second is therefore actually

$$\frac{1}{S} \cdot p_n e^{-\frac{\epsilon_n}{kT}} B_n^m \rho'(\nu', \phi') = \frac{d\kappa'}{4\pi}.$$

In each of these elementary processes the momentum

$$\frac{\epsilon_m - \epsilon_n}{c} \ \cos\phi'$$

is transferred to the molecule in the direction of the X'-axis. In an analogous manner we find, based on (B') that the corresponding number of elementary processes of radiation events of type $Z_m \to Z_n$ per second is

$$\frac{1}{S} \cdot p_m e^{-\frac{\epsilon_m}{kT}} B_m^n \rho'(\nu', \phi') \ \frac{d\kappa'}{4\pi}$$

and in each such elementary process the momentum

$$-\frac{\epsilon_m - \epsilon_n}{c} \ \cos\phi'$$

is transferred to the molecule. The total momentum transferred to the molecule per unit time by incident radiation is, keeping in mind (6) and (9),

$$\frac{h\nu'}{cS} \cdot p_n B_n^m \left(e^{-\frac{\epsilon_n}{kT}} - e^{-\frac{\epsilon_m}{kT}} \right) \int \rho'(\nu', \phi') \cos\phi' \ \frac{d\kappa'}{4\pi}$$

where the integration is to be taken over all solid angles. Carry this out, and we obtain with the aid of (19) value

$$-\frac{h\nu}{c^2 S} \left(\rho - (1/3) \ \nu \ \frac{\partial\rho}{\partial\nu} \right) p_n B_n^m \left(e^{-\frac{\epsilon_n}{kT}} - e^{-\frac{\epsilon_m}{kT}} \right) v.$$

Here we have represented the effective frequency again with ν and not with ν'. This expression gives, however, the total momentum transferred on the average to a molecule moving with speed v. For it is clear that those elementary processes of emission of radiation not induced by the action of the radiation field have no preferred direction as seen from system K' and hence, on the average, cannot transfer any momentum to the molecule. We thus obtain as the final of our considerations

$$R = \frac{h\nu}{c^2 S}\left(\rho - 1/3\ \nu\ \frac{\partial\rho}{\partial\nu}\right) p_n B_n^m e^{-\frac{\epsilon_n}{kT}}\left(1 - e^{-\frac{h\nu}{kT}}\right) \tag{21}$$

Calculating $\overline{\Delta^2}$

It is much easier to calculate the random effect of the elementary processes on the mechanical behavior of the molecule. For we calculate this for a molecule at rest for which the approximation which we have been using applies.

Let some event cause the momentum λ to be transferred to a molecule in the X direction. This momentum is to be of varying magnitude and direction from moment to moment. However, let λ obey a statistical law such that its average value vanishes. Then let $\lambda_1, \lambda_2\ .\ .\ .$ be the momenta which are transferred to the molecule in the X-direction by various operating causes that are independent of each other so that the total momentum that is transferred is

$$\Delta = \Sigma\lambda_\nu$$

We then have (if for the individual $\overline{\lambda_\nu}$ vanish)

$$\overline{\Delta^2} = \overline{\Sigma\lambda_{\nu^2}} \tag{22}$$

If the mean values $\overline{\lambda_{\nu^2}}$ of the individual momenta are all equal to each other $(= \overline{\lambda^2})$ and if l is the total number of processes giving rise to momenta, we have the relation

$$\overline{\Delta^2} = \overline{l\lambda^2} \tag{22a}$$

According to our hypothesis, in each process of incident radiation and outflowing radiation, the momentum

$$\lambda = \frac{h\nu}{c}\cos\phi$$

is transferred to the molecule. Here ϕ is the angle between the X-axis and some randomly chosen direction. Hence, we obtain

$$\overline{\lambda^2} = \frac{1}{3}\left(\frac{h\nu}{c}\right)^2.$$

Since we assume that all the elementary processes that are present are to be considered sas events that are independent of each other, we may apply (22a), l is then the number of all elementary processes that occur in the time τ. This is twice as large as the number of radiation-incident processes $Z_n \to Z_m$ in the time τ. We thus have

$$l = \frac{2}{S} \cdot p_n B_n^m e^{-\frac{\epsilon_n}{kT}} \rho\tau \tag{24}$$

From (23), (24) and (22) we thus obtain

$$\frac{\overline{\Delta^2}}{\tau} = \frac{2}{3S}\left(\frac{h\nu}{c}\right)^2 p_n B_n^m e^{-\frac{\epsilon_n}{kT}} \rho \tag{25}$$

Results

In order now to show that the momenta transferred from the radiation to the molecule according to our basic hypotheses never disturb the thermodynamic equilibrium, we need only introduce the values for $\frac{\overline{\Delta^2}}{\tau}$ and R calculated in (25) and (21) respectively after the quantity

$$\left(\rho - ({}^1/_3)\ \nu\ \frac{\partial\rho}{\partial\nu}\right)\left(1 - e^{-\frac{h\nu}{kT}}\right)$$

in (21) is replaced by

$$\frac{\rho h\nu}{3RT}$$

from (4). We then see that our fundamental equation (12) is satisfied identically.

The above consideration lends very strong support to the hypotheses introduced earlier for the interaction between matter and radiation by means of absorption and emission, and through incident and outgoing radiation. I was led these hypotheses in trying to postulate in the simplest possible way a quantum behavior of molecules that is analogous to the Planck resonators of classical theory. We obtained, without effort, from the general quantum assumption for matter, the second Bohr rule (equation (9)) as well as Planck's radiation formula.

Most important, however, appears to me the result about the momentum transferred to the molecule by incoming and outgoing radiation. If one of our hypotheses were altered, the result would be a violation of equation (12); it appears hardly possible, except by way of our hypotheses, to be in agreement with this relationship which is demanded by thermodynamics. We may therefore consider the following as pretty well proven.

If beam of radiation has the effect that a molecule on which it is incident absorbs or emits an amount of energy h_ν in the from of radiation by means of an elementary process, then the momentum h_ν/c is always transferred to the molecule, and, to be sure, in the case of absorption, in the direction of the moving beam and in the case of emission in the opposite direction. If the molecule is subject to the simultaneous action of beams moving in various directions, then only one of these taken part in any single elementary process of incident radiation; this beam alone then determined the direction of the momentum transferred to the molecule.

If, through an emission process, the molecule suffers a radiant loss of energy of magnitude h_ν without the action of an outside agency, then this process, too, is a directed one. emission in spherical waves does not occur. According to the present state of the theory, the molecule suffers a recoil of magnitude h_ν/c in a particular direction only because of the chance emission in that direction.

This property of elementary processes as expressed by equation (12) makes a quantum theory of radiation almost unavoidable. The weakness of the theory lies, on the one hand, in its not bringing us closer to a union with the wave theory, and, on the other hand, that it leaves the time and direction of the elementary processes to chance; in spite of this, I have full confidence in the trustworthiness of this approach.

Only one more general remark. Almost all theories of thermal radiation rest on the considerations of the interaction between radiation and molecules. But, in general, one is satisfied with dealing only with the energy exchange, without taking into account the momentum exchange. One feels justified in this because the momentum transferred by radiation is so small that it always drops out as compared to that from other dynamical processes. But for the theoretical considerations, this small effect is on an equal footing with the energy transferred by radiation because energy and momentum are very intimately related to each other; a theory may therefore be considered correct only if it can shown that the momentum transferred accordingly from the radiation to the matter leads to the kind of motion that is demanded by thermodynamics.

ON THE ELECTRODYNAMICS OF MOVING BODIES

By A. EINSTEIN

June 30, 1905

It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.

Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good.[1] We will raise this conjecture (the purport of which will hereafter be called the "Principle of Relativity") to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies. The introduction of a "luminiferous ether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space" provided with special properties, nor

[1] The preceding memoir by Lorentz was not at this time known to the author.

assign a velocity-vector to a point of the empty space in which electromagnetic processes take place.

The theory to be developed is based—like all electrodynamics—on the kinematics of the rigid body, since the assertions of any such theory have to do with the relationships between rigid bodies (systems of co-ordinates), clocks, and electromagnetic processes. Insufficient consideration of this circumstance lies at the root of the difficulties which the electrodynamics of moving bodies at present encounters.

I. KINEMATICAL PART

§ 1. Definition of Simultaneity

Let us take a system of co-ordinates in which the equations of Newtonian mechanics hold good.[2] In order to render our presentation more precise and to distinguish this system of co-ordinates verbally from others which will be introduced hereafter, we call it the "stationary system."

If a material point is at rest relatively to this system of co-ordinates, its position can be defined relatively thereto by the employment of rigid standards of measurement and the methods of Euclidean geometry, and can be expressed in Cartesian co-ordinates.

If we wish to describe the *motion* of a material point, we give the values of its co-ordinates as functions of the time. Now we must bear carefully in mind that a mathematical description of this kind has no physical meaning unless we are quite clear as to what we understand by "time." We have to take into account that all our judgments in which time plays a part are always judgments of *simultaneous events*. If, for instance, I say, "That train arrives here at 7 o'clock," I mean something like this: "The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events."[3]

It might appear possible to overcome all the difficulties attending the definition of "time" by substituting "the position of the small hand of my watch" for "time." And in fact such a definition is satisfactory when we are concerned with defining a time exclusively for the place where the watch is located; but it is no longer satisfactory when we have to connect in time series of events occurring at different places, or—what comes to the same thing—to evaluate the times of events occurring at places remote from the watch.

We might, of course, content ourselves with time values determined by an observer stationed together with the watch at the origin of the co-ordinates, and co-ordinating the corresponding positions of the hands with light signals, given out by every event to be timed, and reaching him through empty space. But this co-ordination has the disadvantage that it is not independent of the standpoint of the observer with the watch or clock, as we know from experience.

[2] i.e. to the first approximation.

[3] We shall not here discuss the inexactitude which lurks in the concept of simultaneity of two events at approximately the same place, which can only be removed by an abstraction.

We arrive at a much more practical determination along the following line of thought.

If at the point A of space there is a clock, an observer at A can determine the time values of events in the immediate proximity of A by finding the positions of the hands which are simultaneous with these events. If there is at the point B of space another clock in all respects resembling the one at A, it is possible for an observer at B to determine the time values of events in the immediate neighbourhood of B. But it is not possible without further assumption to compare, in respect of time, an event at A with an event at B. We have so far defined only an "A time" and a "B time." We have not defined a common "time" for A and B, for the latter cannot be defined at all unless we establish *by definition* that the "time" required by light to travel from A to B equals the "time" it requires to travel from B to A. Let a ray of light start at the "A time" t_{A} from A towards B, let it at the "B time" t_{B} be reflected at B in the direction of A, and arrive again at A at the "A time" t'_{A}.

In accordance with definition the two clocks synchronize if

$$t_{\mathrm{B}} - t_{\mathrm{A}} = t'_{\mathrm{A}} - t_{\mathrm{B}}.$$

We assume that this definition of synchronism is free from contradictions, and possible for any number of points; and that the following relations are universally valid:—

1. If the clock at B synchronizes with the clock at A, the clock at A synchronizes with the clock at B.

2. If the clock at A synchronizes with the clock at B and also with the clock at C, the clocks at B and C also synchronize with each other.

Thus with the help of certain imaginary physical experiments we have settled what is to be understood by synchronous stationary clocks located at different places, and have evidently obtained a definition of "simultaneous," or "synchronous," and of "time." The "time" of an event is that which is given simultaneously with the event by a stationary clock located at the place of the event, this clock being synchronous, and indeed synchronous for all time determinations, with a specified stationary clock.

In agreement with experience we further assume the quantity

$$\frac{2\mathrm{AB}}{t'_A - t_A} = c,$$

to be a universal constant—the velocity of light in empty space.

It is essential to have time defined by means of stationary clocks in the stationary system, and the time now defined being appropriate to the stationary system we call it "the time of the stationary system."

§ 2. On the Relativity of Lengths and Times

The following reflexions are based on the principle of relativity and on the principle of the constancy of the velocity of light. These two principles we define as follows:—

1. The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of co-ordinates in uniform translatory motion.

2. Any ray of light moves in the "stationary" system of co-ordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body. Hence

$$\text{velocity} = \frac{\text{light path}}{\text{time interval}}$$

where time interval is to be taken in the sense of the definition in § 1.

Let there be given a stationary rigid rod; and let its length be l as measured by a measuring-rod which is also stationary. We now imagine the axis of the rod lying along the axis of x of the stationary system of co-ordinates, and that a uniform motion of parallel translation with velocity v along the axis of x in the direction of increasing x is then imparted to the rod. We now inquire as to the length of the moving rod, and imagine its length to be ascertained by the following two operations:—

(a) The observer moves together with the given measuring-rod and the rod to be measured, and measures the length of the rod directly by superposing the measuring-rod, in just the same way as if all three were at rest.

(b) By means of stationary clocks set up in the stationary system and synchronizing in accordance with § 1, the observer ascertains at what points of the stationary system the two ends of the rod to be measured are located at a definite time. The distance between these two points, measured by the measuring-rod already employed, which in this case is at rest, is also a length which may be designated "the length of the rod."

In accordance with the principle of relativity the length to be discovered by the operation (a)—we will call it "the length of the rod in the moving system"—must be equal to the length l of the stationary rod.

The length to be discovered by the operation (b) we will call "the length of the (moving) rod in the stationary system." This we shall determine on the basis of our two principles, and we shall find that it differs from l.

Current kinematics tacitly assumes that the lengths determined by these two operations are precisely equal, or in other words, that a moving rigid body at the epoch t may in geometrical respects be perfectly represented by *the same* body *at rest* in a definite position.

We imagine further that at the two ends A and B of the rod, clocks are placed which synchronize with the clocks of the stationary system, that is to say that their indications correspond at any instant to the "time of the stationary system" at the places where they happen to be. These clocks are therefore "synchronous in the stationary system."

We imagine further that with each clock there is a moving observer, and that these observers apply to both clocks the criterion established in § 1 for the synchronization of two clocks. Let a ray of light depart from A at the time[4] $t_{\rm A}$,

[4] "Time" here denotes "time of the stationary system" and also "position of hands of the moving clock situated at the place under discussion."

let it be reflected at B at the time t_{B}, and reach A again at the time t'_{A}. Taking into consideration the principle of the constancy of the velocity of light we find that

$$t_{\mathrm{B}} - t_{\mathrm{A}} = \frac{r_{\mathrm{AB}}}{c - v} \text{ and } t'_{\mathrm{A}} - t_{\mathrm{B}} = \frac{r_{\mathrm{AB}}}{c + v}$$

where r_{AB} denotes the length of the moving rod—measured in the stationary system. Observers moving with the moving rod would thus find that the two clocks were not synchronous, while observers in the stationary system would declare the clocks to be synchronous.

So we see that we cannot attach any *absolute* signification to the concept of simultaneity, but that two events which, viewed from a system of co-ordinates, are simultaneous, can no longer be looked upon as simultaneous events when envisaged from a system which is in motion relatively to that system.

§ 3. Theory of the Transformation of Co-ordinates and Times from a Stationary System to another System in Uniform Motion of Translation Relatively to the Former

Let us in "stationary" space take two systems of co-ordinates, i.e. two systems, each of three rigid material lines, perpendicular to one another, and issuing from a point. Let the axes of X of the two systems coincide, and their axes of Y and Z respectively be parallel. Let each system be provided with a rigid measuring-rod and a number of clocks, and let the two measuring-rods, and likewise all the clocks of the two systems, be in all respects alike.

Now to the origin of one of the two systems (k) let a constant velocity v be imparted in the direction of the increasing x of the other stationary system (K), and let this velocity be communicated to the axes of the co-ordinates, the relevant measuring-rod, and the clocks. To any time of the stationary system K there then will correspond a definite position of the axes of the moving system, and from reasons of symmetry we are entitled to assume that the motion of k may be such that the axes of the moving system are at the time t (this "t" always denotes a time of the stationary system) parallel to the axes of the stationary system.

We now imagine space to be measured from the stationary system K by means of the stationary measuring-rod, and also from the moving system k by means of the measuring-rod moving with it; and that we thus obtain the co-ordinates x, y, z, and ξ, η, ζ respectively. Further, let the time t of the stationary system be determined for all points thereof at which there are clocks by means of light signals in the manner indicated in § 1; similarly let the time τ of the moving system be determined for all points of the moving system at which there are clocks at rest relatively to that system by applying the method, given in § 1, of light signals between the points at which the latter clocks are located.

To any system of values x, y, z, t, which completely defines the place and time of an event in the stationary system, there belongs a system of values ξ,

η, ζ, τ, determining that event relatively to the system k, and our task is now to find the system of equations connecting these quantities.

In the first place it is clear that the equations must be *linear* on account of the properties of homogeneity which we attribute to space and time.

If we place $x' = x - vt$, it is clear that a point at rest in the system k must have a system of values x', y, z, independent of time. We first define τ as a function of x', y, z, and t. To do this we have to express in equations that τ is nothing else than the summary of the data of clocks at rest in system k, which have been synchronized according to the rule given in § 1.

From the origin of system k let a ray be emitted at the time τ_0 along the X-axis to x', and at the time τ_1 be reflected thence to the origin of the co-ordinates, arriving there at the time τ_2; we then must have $\frac{1}{2}(\tau_0 + \tau_2) = \tau_1$, or, by inserting the arguments of the function τ and applying the principle of the constancy of the velocity of light in the stationary system:—

$$\frac{1}{2}\left[\tau(0,0,0,t) + \tau\left(0,0,0,t + \frac{x'}{c-v} + \frac{x'}{c+v}\right)\right] = \tau\left(x',0,0,t + \frac{x'}{c-v}\right).$$

Hence, if x' be chosen infinitesimally small,

$$\frac{1}{2}\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\frac{\partial\tau}{\partial t} = \frac{\partial\tau}{\partial x'} + \frac{1}{c-v}\frac{\partial\tau}{\partial t},$$

or

$$\frac{\partial\tau}{\partial x'} + \frac{v}{c^2 - v^2}\frac{\partial\tau}{\partial t} = 0.$$

It is to be noted that instead of the origin of the co-ordinates we might have chosen any other point for the point of origin of the ray, and the equation just obtained is therefore valid for all values of x', y, z.

An analogous consideration—applied to the axes of Y and Z—it being borne in mind that light is always propagated along these axes, when viewed from the stationary system, with the velocity $\sqrt{c^2 - v^2}$ gives us

$$\frac{\partial\tau}{\partial y} = 0, \frac{\partial\tau}{\partial z} = 0.$$

Since τ is a *linear* function, it follows from these equations that

$$\tau = a\left(t - \frac{v}{c^2 - v^2}x'\right)$$

where a is a function $\phi(v)$ at present unknown, and where for brevity it is assumed that at the origin of k, $\tau = 0$, when $t = 0$.

With the help of this result we easily determine the quantities ξ, η, ζ by expressing in equations that light (as required by the principle of the constancy of the velocity of light, in combination with the principle of relativity) is also propagated with velocity c when measured in the moving system. For a ray of light emitted at the time $\tau = 0$ in the direction of the increasing ξ

$$\xi = c\tau \text{ or } \xi = ac\left(t - \frac{v}{c^2 - v^2}x'\right).$$

But the ray moves relatively to the initial point of k, when measured in the stationary system, with the velocity $c - v$, so that

$$\frac{x'}{c - v} = t.$$

If we insert this value of t in the equation for ξ, we obtain

$$\xi = a\frac{c^2}{c^2 - v^2}x'.$$

In an analogous manner we find, by considering rays moving along the two other axes, that

$$\eta = c\tau = ac\left(t - \frac{v}{c^2 - v^2}x'\right)$$

when

$$\frac{y}{\sqrt{c^2 - v^2}} = t,\ x' = 0.$$

Thus

$$\eta = a\frac{c}{\sqrt{c^2 - v^2}}y \text{ and } \zeta = a\frac{c}{\sqrt{c^2 - v^2}}z.$$

Substituting for x' its value, we obtain

$$\begin{aligned}
\tau &= \phi(v)\beta(t - vx/c^2), \\
\xi &= \phi(v)\beta(x - vt), \\
\eta &= \phi(v)y, \\
\zeta &= \phi(v)z,
\end{aligned}$$

where

$$\beta = \frac{1}{\sqrt{1 - v^2/c^2}},$$

and ϕ is an as yet unknown function of v. If no assumption whatever be made as to the initial position of the moving system and as to the zero point of τ, an additive constant is to be placed on the right side of each of these equations.

We now have to prove that any ray of light, measured in the moving system, is propagated with the velocity c, if, as we have assumed, this is the case in the stationary system; for we have not as yet furnished the proof that the principle of the constancy of the velocity of light is compatible with the principle of relativity.

At the time $t = \tau = 0$, when the origin of the co-ordinates is common to the two systems, let a spherical wave be emitted therefrom, and be propagated with the velocity c in system K. If (x, y, z) be a point just attained by this wave, then

$$x^2 + y^2 + z^2 = c^2 t^2.$$

Transforming this equation with the aid of our equations of transformation we obtain after a simple calculation

$$\xi^2 + \eta^2 + \zeta^2 = c^2 \tau^2.$$

The wave under consideration is therefore no less a spherical wave with velocity of propagation c when viewed in the moving system. This shows that our two fundamental principles are compatible.[5]

In the equations of transformation which have been developed there enters an unknown function ϕ of v, which we will now determine.

For this purpose we introduce a third system of co-ordinates K', which relatively to the system k is in a state of parallel translatory motion parallel to the axis of Ξ,[†] such that the origin of co-ordinates of system K' moves with velocity $-v$ on the axis of Ξ. At the time $t = 0$ let all three origins coincide, and when $t = x = y = z = 0$ let the time t' of the system K' be zero. We call the co-ordinates, measured in the system K', x', y', z', and by a twofold application of our equations of transformation we obtain

$$\begin{array}{lclcl}
t' &=& \phi(-v)\beta(-v)(\tau + v\xi/c^2) &=& \phi(v)\phi(-v)t, \\
x' &=& \phi(-v)\beta(-v)(\xi + v\tau) &=& \phi(v)\phi(-v)x, \\
y' &=& \phi(-v)\eta &=& \phi(v)\phi(-v)y, \\
z' &=& \phi(-v)\zeta &=& \phi(v)\phi(-v)z.
\end{array}$$

Since the relations between x', y', z' and x, y, z do not contain the time t, the systems K and K' are at rest with respect to one another, and it is clear that the transformation from K to K' must be the identical transformation. Thus

$$\phi(v)\phi(-v) = 1.$$

[5] The equations of the Lorentz transformation may be more simply deduced directly from the condition that in virtue of those equations the relation $x^2 + y^2 + z^2 = c^2 t^2$ shall have as its consequence the second relation $\xi^2 + \eta^2 + \zeta^2 = c^2 \tau^2$.

[†] Editor's note: In Einstein's original paper, the symbols $(\Xi, \mathrm{H}, \mathrm{Z})$ for the co-ordinates of the moving system k were introduced without explicitly defining them. In the 1923 English translation, $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ were used, creating an ambiguity between X co-ordinates in the fixed system K and the parallel axis in moving system k. Here and in subsequent references we use Ξ when referring to the axis of system k along which the system is translating with respect to K. In addition, the reference to system K' later in this sentence was incorrectly given as "k" in the 1923 English translation.

We now inquire into the signification of $\phi(v)$. We give our attention to that part of the axis of Y of system k which lies between $\xi = 0, \eta = 0, \zeta = 0$ and $\xi = 0, \eta = l, \zeta = 0$. This part of the axis of Y is a rod moving perpendicularly to its axis with velocity v relatively to system K. Its ends possess in K the co-ordinates

$$x_1 = vt,\ y_1 = \frac{l}{\phi(v)},\ z_1 = 0$$

and

$$x_2 = vt,\ y_2 = 0,\ z_2 = 0.$$

The length of the rod measured in K is therefore $l/\phi(v)$; and this gives us the meaning of the function $\phi(v)$. From reasons of symmetry it is now evident that the length of a given rod moving perpendicularly to its axis, measured in the stationary system, must depend only on the velocity and not on the direction and the sense of the motion. The length of the moving rod measured in the stationary system does not change, therefore, if v and $-v$ are interchanged. Hence follows that $l/\phi(v) = l/\phi(-v)$, or

$$\phi(v) = \phi(-v).$$

It follows from this relation and the one previously found that $\phi(v) = 1$, so that the transformation equations which have been found become

$$\begin{aligned}
\tau &= \beta(t - vx/c^2),\\
\xi &= \beta(x - vt),\\
\eta &= y,\\
\zeta &= z,
\end{aligned}$$

where

$$\beta = 1/\sqrt{1 - v^2/c^2}.$$

§ 4. Physical Meaning of the Equations Obtained in Respect to Moving Rigid Bodies and Moving Clocks

We envisage a rigid sphere[6] of radius R, at rest relatively to the moving system k, and with its centre at the origin of co-ordinates of k. The equation of the surface of this sphere moving relatively to the system K with velocity v is

$$\xi^2 + \eta^2 + \zeta^2 = \mathrm{R}^2.$$

[6] That is, a body possessing spherical form when examined at rest.

The equation of this surface expressed in x, y, z at the time $t = 0$ is

$$\frac{x^2}{(\sqrt{1 - v^2/c^2})^2} + y^2 + z^2 = \mathrm{R}^2.$$

A rigid body which, measured in a state of rest, has the form of a sphere, therefore has in a state of motion—viewed from the stationary system—the form of an ellipsoid of revolution with the axes

$$\mathrm{R}\sqrt{1 - v^2/c^2},\ \mathrm{R},\ \mathrm{R}.$$

Thus, whereas the Y and Z dimensions of the sphere (and therefore of every rigid body of no matter what form) do not appear modified by the motion, the X dimension appears shortened in the ratio $1 : \sqrt{1 - v^2/c^2}$, i.e. the greater the value of v, the greater the shortening. For $v = c$ all moving objects—viewed from the "stationary" system—shrivel up into plane figures.† For velocities greater than that of light our deliberations become meaningless; we shall, however, find in what follows, that the velocity of light in our theory plays the part, physically, of an infinitely great velocity.

It is clear that the same results hold good of bodies at rest in the "stationary" system, viewed from a system in uniform motion.

Further, we imagine one of the clocks which are qualified to mark the time t when at rest relatively to the stationary system, and the time τ when at rest relatively to the moving system, to be located at the origin of the co-ordinates of k, and so adjusted that it marks the time τ. What is the rate of this clock, when viewed from the stationary system?

Between the quantities x, t, and τ, which refer to the position of the clock, we have, evidently, $x = vt$ and

$$\tau = \frac{1}{\sqrt{1 - v^2/c^2}}(t - vx/c^2).$$

Therefore,

$$\tau = t\sqrt{1 - v^2/c^2} = t - (1 - \sqrt{1 - v^2/c^2})t$$

whence it follows that the time marked by the clock (viewed in the stationary system) is slow by $1 - \sqrt{1 - v^2/c^2}$ seconds per second, or—neglecting magnitudes of fourth and higher order—by $\frac{1}{2}v^2/c^2$.

From this there ensues the following peculiar consequence. If at the points A and B of K there are stationary clocks which, viewed in the stationary system, are synchronous; and if the clock at A is moved with the velocity v along the line AB to B, then on its arrival at B the two clocks no longer synchronize, but the clock moved from A to B lags behind the other which has remained at

†Editor's note: In the 1923 English translation, this phrase was erroneously translated as "plain figures". I have used the correct "plane figures" in this edition.

B by $\frac{1}{2}tv^2/c^2$ (up to magnitudes of fourth and higher order), t being the time occupied in the journey from A to B.

It is at once apparent that this result still holds good if the clock moves from A to B in any polygonal line, and also when the points A and B coincide.

If we assume that the result proved for a polygonal line is also valid for a continuously curved line, we arrive at this result: If one of two synchronous clocks at A is moved in a closed curve with constant velocity until it returns to A, the journey lasting t seconds, then by the clock which has remained at rest the travelled clock on its arrival at A will be $\frac{1}{2}tv^2/c^2$ second slow. Thence we conclude that a balance-clock[7] at the equator must go more slowly, by a very small amount, than a precisely similar clock situated at one of the poles under otherwise identical conditions.

§ 5. The Composition of Velocities

In the system k moving along the axis of X of the system K with velocity v, let a point move in accordance with the equations

$$\xi = w_\xi \tau, \eta = w_\eta \tau, \zeta = 0,$$

where w_ξ and w_η denote constants.

Required: the motion of the point relatively to the system K. If with the help of the equations of transformation developed in § 3 we introduce the quantities x, y, z, t into the equations of motion of the point, we obtain

$$\begin{aligned} x &= \frac{w_\xi + v}{1 + v w_\xi / c^2} t, \\ y &= \frac{\sqrt{1 - v^2/c^2}}{1 + v w_\xi / c^2} w_\eta t, \\ z &= 0. \end{aligned}$$

Thus the law of the parallelogram of velocities is valid according to our theory only to a first approximation. We set

$$\begin{aligned} V^2 &= \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2, \\ w^2 &= w_\xi^2 + w_\eta^2, \\ a &= \tan^{-1} w_\eta / w_\xi, ^\dagger \end{aligned}$$

[7] Not a pendulum-clock, which is physically a system to which the Earth belongs. This case had to be excluded.

† Editor's note: This equation was incorrectly given in Einstein's original paper and the 1923 English translation as $a = \tan^{-1} w_y / w_x$.

a is then to be looked upon as the angle between the velocities v and w. After a simple calculation we obtain

$$V = \frac{\sqrt{(v^2 + w^2 + 2vw\cos a) - (vw\sin a/c)^2}}{1 + vw\cos a/c^2}.$$

It is worthy of remark that v and w enter into the expression for the resultant velocity in a symmetrical manner. If w also has the direction of the axis of X, we get

$$V = \frac{v + w}{1 + vw/c^2}.$$

It follows from this equation that from a composition of two velocities which are less than c, there always results a velocity less than c. For if we set $v = c - \kappa, w = c - \lambda$, κ and λ being positive and less than c, then

$$V = c\frac{2c - \kappa - \lambda}{2c - \kappa - \lambda + \kappa\lambda/c} < c.$$

It follows, further, that the velocity of light c cannot be altered by composition with a velocity less than that of light. For this case we obtain

$$V = \frac{c + w}{1 + w/c} = c.$$

We might also have obtained the formula for V, for the case when v and w have the same direction, by compounding two transformations in accordance with § 3. If in addition to the systems K and k figuring in § 3 we introduce still another system of co-ordinates k' moving parallel to k, its initial point moving on the axis of Ξ† with the velocity w, we obtain equations between the quantities x, y, z, t and the corresponding quantities of k', which differ from the equations found in § 3 only in that the place of "v" is taken by the quantity

$$\frac{v + w}{1 + vw/c^2};$$

from which we see that such parallel transformations—necessarily—form a group.

We have now deduced the requisite laws of the theory of kinematics corresponding to our two principles, and we proceed to show their application to electrodynamics.

II. ELECTRODYNAMICAL PART

§ 6. Transformation of the Maxwell-Hertz Equations for Empty Space. On the Nature of the Electromotive Forces Occurring in a Magnetic Field During Motion

Let the Maxwell-Hertz equations for empty space hold good for the stationary system K, so that we have

†Editor's note: "X" in the 1923 English translation.

$$
\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}}{\partial t} &= \frac{\partial \mathrm{N}}{\partial y} - \frac{\partial \mathrm{M}}{\partial z}, & \frac{1}{c}\frac{\partial \mathrm{L}}{\partial t} &= \frac{\partial \mathrm{Y}}{\partial z} - \frac{\partial \mathrm{Z}}{\partial y}, \\
\frac{1}{c}\frac{\partial \mathrm{Y}}{\partial t} &= \frac{\partial \mathrm{L}}{\partial z} - \frac{\partial \mathrm{N}}{\partial x}, & \frac{1}{c}\frac{\partial \mathrm{M}}{\partial t} &= \frac{\partial \mathrm{Z}}{\partial x} - \frac{\partial \mathrm{X}}{\partial z}, \\
\frac{1}{c}\frac{\partial \mathrm{Z}}{\partial t} &= \frac{\partial \mathrm{M}}{\partial x} - \frac{\partial \mathrm{L}}{\partial y}, & \frac{1}{c}\frac{\partial \mathrm{N}}{\partial t} &= \frac{\partial \mathrm{X}}{\partial y} - \frac{\partial \mathrm{Y}}{\partial x},
\end{aligned}
$$

where (X, Y, Z) denotes the vector of the electric force, and (L, M, N) that of the magnetic force.

If we apply to these equations the transformation developed in § 3, by referring the electromagnetic processes to the system of co-ordinates there introduced, moving with the velocity v, we obtain the equations †

$$
\begin{aligned}
\tfrac{1}{c}\tfrac{\partial \mathrm{X}}{\partial \tau} &= \tfrac{\partial}{\partial \eta}\left\{\beta\left(\mathrm{N} - \tfrac{v}{c}\mathrm{Y}\right)\right\} &&- \tfrac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{M} + \tfrac{v}{c}\mathrm{Z}\right)\right\}, \\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{Y} - \tfrac{v}{c}\mathrm{N}\right)\right\} &= \tfrac{\partial \mathrm{L}}{\partial \zeta} &&- \tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{N} - \tfrac{v}{c}\mathrm{Y}\right)\right\}, \\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{Z} + \tfrac{v}{c}\mathrm{M}\right)\right\} &= \tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{M} + \tfrac{v}{c}\mathrm{Z}\right)\right\} &&- \tfrac{\partial \mathrm{L}}{\partial \eta}, \\
\tfrac{1}{c}\tfrac{\partial \mathrm{L}}{\partial \tau} &= \tfrac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{Y} - \tfrac{v}{c}\mathrm{N}\right)\right\} &&- \tfrac{\partial}{\partial \eta}\left\{\beta\left(\mathrm{Z} + \tfrac{v}{c}\mathrm{M}\right)\right\}, \\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{M} + \tfrac{v}{c}\mathrm{Z}\right)\right\} &= \tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{Z} + \tfrac{v}{c}\mathrm{M}\right)\right\} &&- \tfrac{\partial \mathrm{X}}{\partial \zeta}, \\
\tfrac{1}{c}\tfrac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{N} - \tfrac{v}{c}\mathrm{Y}\right)\right\} &= \tfrac{\partial \mathrm{X}}{\partial \eta} &&- \tfrac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{Y} - \tfrac{v}{c}\mathrm{N}\right)\right\},
\end{aligned}
$$

where

$$\beta = 1/\sqrt{1 - v^2/c^2}.$$

Now the principle of relativity requires that if the Maxwell-Hertz equations for empty space hold good in system K, they also hold good in system k; that is to say that the vectors of the electric and the magnetic force—(X′, Y′, Z′) and (L′, M′, N′)—of the moving system k, which are defined by their ponderomotive effects on electric or magnetic masses respectively, satisfy the following equations:—

$$
\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}'}{\partial \tau} &= \frac{\partial \mathrm{N}'}{\partial \eta} - \frac{\partial \mathrm{M}'}{\partial \zeta}, & \frac{1}{c}\frac{\partial \mathrm{L}'}{\partial \tau} &= \frac{\partial \mathrm{Y}'}{\partial \zeta} - \frac{\partial \mathrm{Z}'}{\partial \eta}, \\
\frac{1}{c}\frac{\partial \mathrm{Y}'}{\partial \tau} &= \frac{\partial \mathrm{L}'}{\partial \zeta} - \frac{\partial \mathrm{N}'}{\partial \xi}, & \frac{1}{c}\frac{\partial \mathrm{M}'}{\partial \tau} &= \frac{\partial \mathrm{Z}'}{\partial \xi} - \frac{\partial \mathrm{X}'}{\partial \zeta}, \\
\frac{1}{c}\frac{\partial \mathrm{Z}'}{\partial \tau} &= \frac{\partial \mathrm{M}'}{\partial \xi} - \frac{\partial \mathrm{L}'}{\partial \eta}, & \frac{1}{c}\frac{\partial \mathrm{N}'}{\partial \tau} &= \frac{\partial \mathrm{X}'}{\partial \eta} - \frac{\partial \mathrm{Y}'}{\partial \xi}.
\end{aligned}
$$

† Editor's note: In the 1923 English translation, the quantities "ζ" and "ξ" were interchanged in the second equation. They were given correctly in the the original 1905 paper.

Evidently the two systems of equations found for system k must express exactly the same thing, since both systems of equations are equivalent to the Maxwell-Hertz equations for system K. Since, further, the equations of the two systems agree, with the exception of the symbols for the vectors, it follows that the functions occurring in the systems of equations at corresponding places must agree, with the exception of a factor $\psi(v)$, which is common for all functions of the one system of equations, and is independent of ξ, η, ζ and τ but depends upon v. Thus we have the relations

$$\begin{array}{lll} \mathrm{X}' = \psi(v)\mathrm{X}, & \mathrm{L}' = \psi(v)\mathrm{L}, \\ \mathrm{Y}' = \psi(v)\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right), & \mathrm{M}' = \psi(v)\beta\left(\mathrm{M} + \frac{v}{c}\mathrm{Z}\right), \\ \mathrm{Z}' = \psi(v)\beta\left(\mathrm{Z} + \frac{v}{c}\mathrm{M}\right), & \mathrm{N}' = \psi(v)\beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right). \end{array}$$

If we now form the reciprocal of this system of equations, firstly by solving the equations just obtained, and secondly by applying the equations to the inverse transformation (from k to K), which is characterized by the velocity $-v$, it follows, when we consider that the two systems of equations thus obtained must be identical, that $\psi(v)\psi(-v) = 1$. Further, from reasons of symmetry[8] and therefore

$$\psi(v) = 1,$$

and our equations assume the form

$$\begin{array}{lll} \mathrm{X}' = \mathrm{X}, & \mathrm{L}' = \mathrm{L}, \\ \mathrm{Y}' = \beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right), & \mathrm{M}' = \beta\left(\mathrm{M} + \frac{v}{c}\mathrm{Z}\right), \\ \mathrm{Z}' = \beta\left(\mathrm{Z} + \frac{v}{c}\mathrm{M}\right), & \mathrm{N}' = \beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right). \end{array}$$

As to the interpretation of these equations we make the following remarks: Let a point charge of electricity have the magnitude "one" when measured in the stationary system K, i.e. let it when at rest in the stationary system exert a force of one dyne upon an equal quantity of electricity at a distance of one cm. By the principle of relativity this electric charge is also of the magnitude "one" when measured in the moving system. If this quantity of electricity is at rest relatively to the stationary system, then by definition the vector (X, Y, Z) is equal to the force acting upon it. If the quantity of electricity is at rest relatively to the moving system (at least at the relevant instant), then the force acting upon it, measured in the moving system, is equal to the vector (X′, Y′, Z′). Consequently the first three equations above allow themselves to be clothed in words in the two following ways:—

1. If a unit electric point charge is in motion in an electromagnetic field, there acts upon it, in addition to the electric force, an "electromotive force" which, if we neglect the terms multiplied by the second and higher powers of

[8] If, for example, X=Y=Z=L=M=0, and N $\neq 0$, then from reasons of symmetry it is clear that when v changes sign without changing its numerical value, Y′ must also change sign without changing its numerical value.

v/c, is equal to the vector-product of the velocity of the charge and the magnetic force, divided by the velocity of light. (Old manner of expression.)

2. If a unit electric point charge is in motion in an electromagnetic field, the force acting upon it is equal to the electric force which is present at the locality of the charge, and which we ascertain by transformation of the field to a system of co-ordinates at rest relatively to the electrical charge. (New manner of expression.)

The analogy holds with "magnetomotive forces." We see that electromotive force plays in the developed theory merely the part of an auxiliary concept, which owes its introduction to the circumstance that electric and magnetic forces do not exist independently of the state of motion of the system of co-ordinates.

Furthermore it is clear that the asymmetry mentioned in the introduction as arising when we consider the currents produced by the relative motion of a magnet and a conductor, now disappears. Moreover, questions as to the "seat" of electrodynamic electromotive forces (unipolar machines) now have no point.

§ 7. Theory of Doppler's Principle and of Aberration

In the system K, very far from the origin of co-ordinates, let there be a source of electrodynamic waves, which in a part of space containing the origin of co-ordinates may be represented to a sufficient degree of approximation by the equations

$$\begin{array}{ll} \mathrm{X} = \mathrm{X}_0 \sin\Phi, & \mathrm{L} = \mathrm{L}_0 \sin\Phi, \\ \mathrm{Y} = \mathrm{Y}_0 \sin\Phi, & \mathrm{M} = \mathrm{M}_0 \sin\Phi, \\ \mathrm{Z} = \mathrm{Z}_0 \sin\Phi, & \mathrm{N} = \mathrm{N}_0 \sin\Phi, \end{array}$$

where

$$\Phi = \omega \left\{ t - \frac{1}{c}(lx + my + nz) \right\}.$$

Here $(\mathrm{X}_0, \mathrm{Y}_0, \mathrm{Z}_0)$ and $(\mathrm{L}_0, \mathrm{M}_0, \mathrm{N}_0)$ are the vectors defining the amplitude of the wave-train, and l, m, n the direction-cosines of the wave-normals. We wish to know the constitution of these waves, when they are examined by an observer at rest in the moving system k.

Applying the equations of transformation found in § 6 for electric and magnetic forces, and those found in § 3 for the co-ordinates and the time, we obtain directly

$$\begin{array}{ll} \mathrm{X}' = \mathrm{X}_0 \sin\Phi', & \mathrm{L}' = \mathrm{L}_0 \sin\Phi', \\ \mathrm{Y}' = \beta(\mathrm{Y}_0 - v\mathrm{N}_0/c)\sin\Phi', & \mathrm{M}' = \beta(\mathrm{M}_0 + v\mathrm{Z}_0/c)\sin\Phi', \\ \mathrm{Z}' = \beta(\mathrm{Z}_0 + v\mathrm{M}_0/c)\sin\Phi', & \mathrm{N}' = \beta(\mathrm{N}_0 - v\mathrm{Y}_0/c)\sin\Phi', \end{array}$$
$$\Phi' = \omega'\left\{\tau - \tfrac{1}{c}(l'\xi + m'\eta + n'\zeta)\right\}$$

where

$$\begin{aligned}
\omega' &= \omega\beta(1 - lv/c), \\
l' &= \frac{l - v/c}{1 - lv/c}, \\
m' &= \frac{m}{\beta(1 - lv/c)}, \\
n' &= \frac{n}{\beta(1 - lv/c)}.
\end{aligned}$$

From the equation for ω' it follows that if an observer is moving with velocity v relatively to an infinitely distant source of light of frequency ν, in such a way that the connecting line "source-observer" makes the angle ϕ with the velocity of the observer referred to a system of co-ordinates which is at rest relatively to the source of light, the frequency ν' of the light perceived by the observer is given by the equation

$$\nu' = \nu \frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}}.$$

This is Doppler's principle for any velocities whatever. When $\phi = 0$ the equation assumes the perspicuous form

$$\nu' = \nu\sqrt{\frac{1 - v/c}{1 + v/c}}.$$

We see that, in contrast with the customary view, when $v = -c, \nu' = \infty$.

If we call the angle between the wave-normal (direction of the ray) in the moving system and the connecting line "source-observer" ϕ', the equation for ϕ'† assumes the form

$$\cos\phi' = \frac{\cos\phi - v/c}{1 - \cos\phi \cdot v/c}.$$

This equation expresses the law of aberration in its most general form. If $\phi = \frac{1}{2}\pi$, the equation becomes simply

$$\cos\phi' = -v/c.$$

We still have to find the amplitude of the waves, as it appears in the moving system. If we call the amplitude of the electric or magnetic force A or A′ respectively, accordingly as it is measured in the stationary system or in the moving system, we obtain

†Editor's note: Erroneously given as "l'" in the 1923 English translation, propagating an error, despite a change in symbols, from the original 1905 paper.

$$\mathrm{A}'^2 = \mathrm{A}^2 \frac{(1 - \cos\phi \cdot v/c)^2}{1 - v^2/c^2}$$

which equation, if $\phi = 0$, simplifies into

$$\mathrm{A}'^2 = \mathrm{A}^2 \frac{1 - v/c}{1 + v/c}.$$

It follows from these results that to an observer approaching a source of light with the velocity c, this source of light must appear of infinite intensity.

§ 8. Transformation of the Energy of Light Rays. Theory of the Pressure of Radiation Exerted on Perfect Reflectors

Since $\mathrm{A}^2/8\pi$ equals the energy of light per unit of volume, we have to regard $\mathrm{A}'^2/8\pi$, by the principle of relativity, as the energy of light in the moving system. Thus $\mathrm{A}'^2/\mathrm{A}^2$ would be the ratio of the "measured in motion" to the "measured at rest" energy of a given light complex, if the volume of a light complex were the same, whether measured in K or in k. But this is not the case. If l, m, n are the direction-cosines of the wave-normals of the light in the stationary system, no energy passes through the surface elements of a spherical surface moving with the velocity of light:—

$$(x - lct)^2 + (y - mct)^2 + (z - nct)^2 = \mathrm{R}^2.$$

We may therefore say that this surface permanently encloses the same light complex. We inquire as to the quantity of energy enclosed by this surface, viewed in system k, that is, as to the energy of the light complex relatively to the system k.

The spherical surface—viewed in the moving system—is an ellipsoidal surface, the equation for which, at the time $\tau = 0$, is

$$(\beta\xi - l\beta\xi v/c)^2 + (\eta - m\beta\xi v/c)^2 + (\zeta - n\beta\xi v/c)^2 = \mathrm{R}^2.$$

If S is the volume of the sphere, and S′ that of this ellipsoid, then by a simple calculation

$$\frac{\mathrm{S}'}{\mathrm{S}} = \frac{\sqrt{1 - v^2/c^2}}{1 - \cos\phi \cdot v/c}.$$

Thus, if we call the light energy enclosed by this surface E when it is measured in the stationary system, and E′ when measured in the moving system, we obtain

$$\frac{\mathrm{E}'}{\mathrm{E}} = \frac{\mathrm{A}'^2\mathrm{S}'}{\mathrm{A}^2\mathrm{S}} = \frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}},$$

and this formula, when $\phi = 0$, simplifies into

$$\frac{\mathrm{E}'}{\mathrm{E}} = \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

It is remarkable that the energy and the frequency of a light complex vary with the state of motion of the observer in accordance with the same law.

Now let the co-ordinate plane $\xi = 0$ be a perfectly reflecting surface, at which the plane waves considered in § 7 are reflected. We seek for the pressure of light exerted on the reflecting surface, and for the direction, frequency, and intensity of the light after reflexion.

Let the incidental light be defined by the quantities A, $\cos\phi$, ν (referred to system K). Viewed from k the corresponding quantities are

$$\begin{aligned}
\mathrm{A}' &= \mathrm{A}\frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}}, \\
\cos\phi' &= \frac{\cos\phi - v/c}{1 - \cos\phi \cdot v/c}, \\
\nu' &= \nu\frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}}.
\end{aligned}$$

For the reflected light, referring the process to system k, we obtain

$$\begin{aligned}
\mathrm{A}'' &= \mathrm{A}' \\
\cos\phi'' &= -\cos\phi' \\
\nu'' &= \nu'
\end{aligned}$$

Finally, by transforming back to the stationary system K, we obtain for the reflected light

$$\begin{aligned}
\mathrm{A}''' &= \mathrm{A}''\frac{1 + cos\phi'' \cdot v/c}{\sqrt{1 - v^2/c^2}} = \mathrm{A}\frac{1 - 2\cos\phi \cdot v/c + v^2/c^2}{1 - v^2/c^2}, \\
\cos\phi''' &= \frac{\cos\phi'' + v/c}{1 + \cos\phi'' \cdot v/c} = -\frac{(1 + v^2/c^2)\cos\phi - 2v/c}{1 - 2\cos\phi \cdot v/c + v^2/c^2}, \\
\nu''' &= \nu''\frac{1 + \cos\phi'' \cdot v/c}{\sqrt{1 - v^2/c^2}} = \nu\frac{1 - 2\cos\phi \cdot v/c + v^2/c^2}{1 - v^2/c^2}.
\end{aligned}$$

The energy (measured in the stationary system) which is incident upon unit area of the mirror in unit time is evidently $\mathrm{A}^2(c\cos\phi - v)/8\pi$. The energy leaving the unit of surface of the mirror in the unit of time is $\mathrm{A}'''^2(-c\cos\phi''' + v)/8\pi$.

The difference of these two expressions is, by the principle of energy, the work done by the pressure of light in the unit of time. If we set down this work as equal to the product $\mathrm{P}v$, where P is the pressure of light, we obtain

$$\mathrm{P} = 2 \cdot \frac{\mathrm{A}^2}{8\pi} \frac{(\cos\phi - v/c)^2}{1 - v^2/c^2}.$$

In agreement with experiment and with other theories, we obtain to a first approximation

$$\mathrm{P} = 2 \cdot \frac{\mathrm{A}^2}{8\pi} \cos^2\phi.$$

All problems in the optics of moving bodies can be solved by the method here employed. What is essential is, that the electric and magnetic force of the light which is influenced by a moving body, be transformed into a system of co-ordinates at rest relatively to the body. By this means all problems in the optics of moving bodies will be reduced to a series of problems in the optics of stationary bodies.

§ 9. Transformation of the Maxwell-Hertz Equations when Convection-Currents are Taken into Account

We start from the equations

$$\begin{aligned}
\frac{1}{c}\left\{\frac{\partial \mathrm{X}}{\partial t} + u_x\rho\right\} &= \frac{\partial \mathrm{N}}{\partial y} - \frac{\partial \mathrm{M}}{\partial z}, & \frac{1}{c}\frac{\partial \mathrm{L}}{\partial t} &= \frac{\partial \mathrm{Y}}{\partial z} - \frac{\partial \mathrm{Z}}{\partial y}, \\
\frac{1}{c}\left\{\frac{\partial \mathrm{Y}}{\partial t} + u_y\rho\right\} &= \frac{\partial \mathrm{L}}{\partial z} - \frac{\partial \mathrm{N}}{\partial x}, & \frac{1}{c}\frac{\partial \mathrm{M}}{\partial t} &= \frac{\partial \mathrm{Z}}{\partial x} - \frac{\partial \mathrm{X}}{\partial z}, \\
\frac{1}{c}\left\{\frac{\partial \mathrm{Z}}{\partial t} + u_z\rho\right\} &= \frac{\partial \mathrm{M}}{\partial x} - \frac{\partial \mathrm{L}}{\partial y}, & \frac{1}{c}\frac{\partial \mathrm{N}}{\partial t} &= \frac{\partial \mathrm{X}}{\partial y} - \frac{\partial \mathrm{Y}}{\partial x},
\end{aligned}$$

where

$$\rho = \frac{\partial \mathrm{X}}{\partial x} + \frac{\partial \mathrm{Y}}{\partial y} + \frac{\partial \mathrm{Z}}{\partial z}$$

denotes 4π times the density of electricity, and (u_x, u_y, u_z) the velocity-vector of the charge. If we imagine the electric charges to be invariably coupled to small rigid bodies (ions, electrons), these equations are the electromagnetic basis of the Lorentzian electrodynamics and optics of moving bodies.

Let these equations be valid in the system K, and transform them, with the assistance of the equations of transformation given in §§ 3 and 6, to the system k. We then obtain the equations

$$
\begin{aligned}
\frac{1}{c}\left\{\frac{\partial \mathrm{X}'}{\partial \tau}+u_\xi\rho'\right\} &= \frac{\partial \mathrm{N}'}{\partial \eta}-\frac{\partial \mathrm{M}'}{\partial \zeta}, & \frac{1}{c}\frac{\partial \mathrm{L}'}{\partial \tau} &= \frac{\partial \mathrm{Y}'}{\partial \zeta}-\frac{\partial \mathrm{Z}'}{\partial \eta},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Y}'}{\partial \tau}+u_\eta\rho'\right\} &= \frac{\partial \mathrm{L}'}{\partial \zeta}-\frac{\partial \mathrm{N}'}{\partial \xi}, & \frac{1}{c}\frac{\partial \mathrm{M}'}{\partial \tau} &= \frac{\partial \mathrm{Z}'}{\partial \xi}-\frac{\partial \mathrm{X}'}{\partial \zeta},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Z}'}{\partial \tau}+u_\zeta\rho'\right\} &= \frac{\partial \mathrm{M}'}{\partial \xi}-\frac{\partial \mathrm{L}'}{\partial \eta}, & \frac{1}{c}\frac{\partial \mathrm{N}'}{\partial \tau} &= \frac{\partial \mathrm{X}'}{\partial \eta}-\frac{\partial \mathrm{Y}'}{\partial \xi},
\end{aligned}
$$

where

$$
\begin{aligned}
u_\xi &= \frac{u_x-v}{1-u_xv/c^2}\\
u_\eta &= \frac{u_y}{\beta(1-u_xv/c^2)}\\
u_\zeta &= \frac{u_z}{\beta(1-u_xv/c^2)},
\end{aligned}
$$

and

$$
\begin{aligned}
\rho' &= \frac{\partial \mathrm{X}'}{\partial \xi}+\frac{\partial \mathrm{Y}'}{\partial \eta}+\frac{\partial Z'}{\partial \zeta}\\
&= \beta(1-u_xv/c^2)\rho.
\end{aligned}
$$

Since—as follows from the theorem of addition of velocities (§ 5)—the vector (u_ξ, u_η, u_ζ) is nothing else than the velocity of the electric charge, measured in the system k, we have the proof that, on the basis of our kinematical principles, the electrodynamic foundation of Lorentz's theory of the electrodynamics of moving bodies is in agreement with the principle of relativity.

In addition I may briefly remark that the following important law may easily be deduced from the developed equations: If an electrically charged body is in motion anywhere in space without altering its charge when regarded from a system of co-ordinates moving with the body, its charge also remains—when regarded from the "stationary" system K—constant.

§ 10. Dynamics of the Slowly Accelerated Electron

Let there be in motion in an electromagnetic field an electrically charged particle (in the sequel called an "electron"), for the law of motion of which we assume as follows:—

If the electron is at rest at a given epoch, the motion of the electron ensues in the next instant of time according to the equations

$$
\begin{aligned}
m\frac{d^2x}{dt^2} &= \epsilon \mathrm{X} \\
m\frac{d^2y}{dt^2} &= \epsilon \mathrm{Y} \\
m\frac{d^2z}{dt^2} &= \epsilon \mathrm{Z}
\end{aligned}
$$

where x, y, z denote the co-ordinates of the electron, and m the mass of the electron, as long as its motion is slow.

Now, secondly, let the velocity of the electron at a given epoch be v. We seek the law of motion of the electron in the immediately ensuing instants of time.

Without affecting the general character of our considerations, we may and will assume that the electron, at the moment when we give it our attention, is at the origin of the co-ordinates, and moves with the velocity v along the axis of X of the system K. It is then clear that at the given moment ($t = 0$) the electron is at rest relatively to a system of co-ordinates which is in parallel motion with velocity v along the axis of X.

From the above assumption, in combination with the principle of relativity, it is clear that in the immediately ensuing time (for small values of t) the electron, viewed from the system k, moves in accordance with the equations

$$
\begin{aligned}
m\frac{d^2\xi}{d\tau^2} &= \epsilon \mathrm{X}', \\
m\frac{d^2\eta}{d\tau^2} &= \epsilon \mathrm{Y}', \\
m\frac{d^2\zeta}{d\tau^2} &= \epsilon \mathrm{Z}',
\end{aligned}
$$

in which the symbols ξ, η, ζ, X', Y', Z' refer to the system k. If, further, we decide that when $t = x = y = z = 0$ then $\tau = \xi = \eta = \zeta = 0$, the transformation equations of §§ 3 and 6 hold good, so that we have

$$
\begin{gathered}
\xi = \beta(x - vt), \eta = y, \zeta = z, \tau = \beta(t - vx/c^2), \\
\mathrm{X}' = \mathrm{X}, \mathrm{Y}' = \beta(\mathrm{Y} - v\mathrm{N}/c), \mathrm{Z}' = \beta(\mathrm{Z} + v\mathrm{M}/c).
\end{gathered}
$$

With the help of these equations we transform the above equations of motion from system k to system K, and obtain

$$
\left.
\begin{aligned}
\tfrac{d^2x}{dt^2} &= \tfrac{\epsilon}{m\beta^3}\mathrm{X} \\
\tfrac{d^2y}{dt^2} &= \tfrac{\epsilon}{m\beta}\left(\mathrm{Y} - \tfrac{v}{c}\mathrm{N}\right) \\
\tfrac{d^2z}{dt^2} &= \tfrac{\epsilon}{m\beta}\left(\mathrm{Z} + \tfrac{v}{c}\mathrm{M}\right)
\end{aligned}
\right\} \quad . \quad . \quad . \qquad \text{(A)}
$$

Taking the ordinary point of view we now inquire as to the "longitudinal" and the "transverse" mass of the moving electron. We write the equations (A) in the form

$$\begin{array}{rclcl} m\beta^3 \frac{d^2x}{dt^2} & = & \epsilon \mathrm{X} & = & \epsilon \mathrm{X}', \\ m\beta^2 \frac{d^2y}{dt^2} & = & \epsilon\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right) & = & \epsilon \mathrm{Y}', \\ m\beta^2 \frac{d^2z}{dt^2} & = & \epsilon\beta\left(\mathrm{Z} + \frac{v}{c}\mathrm{M}\right) & = & \epsilon \mathrm{Z}', \end{array}$$

and remark firstly that $\epsilon\mathrm{X}'$, $\epsilon\mathrm{Y}'$, $\epsilon\mathrm{Z}'$ are the components of the ponderomotive force acting upon the electron, and are so indeed as viewed in a system moving at the moment with the electron, with the same velocity as the electron. (This force might be measured, for example, by a spring balance at rest in the last-mentioned system.) Now if we call this force simply "the force acting upon the electron,"[9] and maintain the equation—mass × acceleration = force—and if we also decide that the accelerations are to be measured in the stationary system K, we derive from the above equations

$$\begin{array}{rcl} \text{Longitudinal mass} & = & \dfrac{m}{(\sqrt{1 - v^2/c^2})^3}. \\ \text{Transverse mass} & = & \dfrac{m}{1 - v^2/c^2}. \end{array}$$

With a different definition of force and acceleration we should naturally obtain other values for the masses. This shows us that in comparing different theories of the motion of the electron we must proceed very cautiously.

We remark that these results as to the mass are also valid for ponderable material points, because a ponderable material point can be made into an electron (in our sense of the word) by the addition of an electric charge, *no matter how small.*

We will now determine the kinetic energy of the electron. If an electron moves from rest at the origin of co-ordinates of the system K along the axis of X under the action of an electrostatic force X, it is clear that the energy withdrawn from the electrostatic field has the value $\int \epsilon \mathrm{X}\, dx$. As the electron is to be slowly accelerated, and consequently may not give off any energy in the form of radiation, the energy withdrawn from the electrostatic field must be put down as equal to the energy of motion W of the electron. Bearing in mind that during the whole process of motion which we are considering, the first of the equations (A) applies, we therefore obtain

$$\mathrm{W} \quad = \quad \int \epsilon \mathrm{X}\, dx = m \int_0^v \beta^3 v\, dv$$

[9]The definition of force here given is not advantageous, as was first shown by M. Planck. It is more to the point to define force in such a way that the laws of momentum and energy assume the simplest form.

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

Thus, when $v = c$, W becomes infinite. Velocities greater than that of light have—as in our previous results—no possibility of existence.

This expression for the kinetic energy must also, by virtue of the argument stated above, apply to ponderable masses as well.

We will now enumerate the properties of the motion of the electron which result from the system of equations (A), and are accessible to experiment.

1. From the second equation of the system (A) it follows that an electric force Y and a magnetic force N have an equally strong deflective action on an electron moving with the velocity v, when $\mathrm{Y} = \mathrm{N}v/c$. Thus we see that it is possible by our theory to determine the velocity of the electron from the ratio of the magnetic power of deflexion A_m to the electric power of deflexion A_e, for any velocity, by applying the law

$$\frac{\mathrm{A}_m}{\mathrm{A}_e} = \frac{v}{c}.$$

This relationship may be tested experimentally, since the velocity of the electron can be directly measured, e.g. by means of rapidly oscillating electric and magnetic fields.

2. From the deduction for the kinetic energy of the electron it follows that between the potential difference, P, traversed and the acquired velocity v of the electron there must be the relationship

$$\mathrm{P} = \int \mathrm{X} dx = \frac{m}{\epsilon} c^2 \left\{ \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right\}.$$

3. We calculate the radius of curvature of the path of the electron when a magnetic force N is present (as the only deflective force), acting perpendicularly to the velocity of the electron. From the second of the equations (A) we obtain

$$-\frac{d^2 y}{dt^2} = \frac{v^2}{\mathrm{R}} = \frac{\epsilon}{m} \frac{v}{c} \mathrm{N} \sqrt{1 - \frac{v^2}{c^2}}$$

or

$$\mathrm{R} = \frac{mc^2}{\epsilon} \cdot \frac{v/c}{\sqrt{1 - v^2/c^2}} \cdot \frac{1}{\mathrm{N}}.$$

These three relationships are a complete expression for the laws according to which, by the theory here advanced, the electron must move.

In conclusion I wish to say that in working at the problem here dealt with I have had the loyal assistance of my friend and colleague M. Besso, and that I am indebted to him for several valuable suggestions.

About this Document

This edition of Einstein's *On the Electrodynamics of Moving Bodies* is based on the English translation of his original 1905 German-language paper (published as *Zur Elektrodynamik bewegter Körper*, in *Annalen der Physik*. **17**:891, 1905) which appeared in the book *The Principle of Relativity*, published in 1923 by Methuen and Company, Ltd. of London. Most of the papers in that collection are English translations from the German *Das Relativatsprinzip*, 4th ed., published by in 1922 by Tuebner. All of these sources are now in the public domain; this document, derived from them, remains in the public domain and may be reproduced in any manner or medium without permission, restriction, attribution, or compensation.

Numbered footnotes are as they appeared in the 1923 edition; editor's notes are marked by a dagger (†) and appear in sans serif type. The 1923 English translation modified the notation used in Einstein's 1905 paper to conform to that in use by the 1920's; for example, c denotes the speed of light, as opposed the V used by Einstein in 1905.

This edition was prepared by John Walker. The current version of this document is available in a variety of formats from the editor's Web site:

`http://www.fourmilab.ch/`

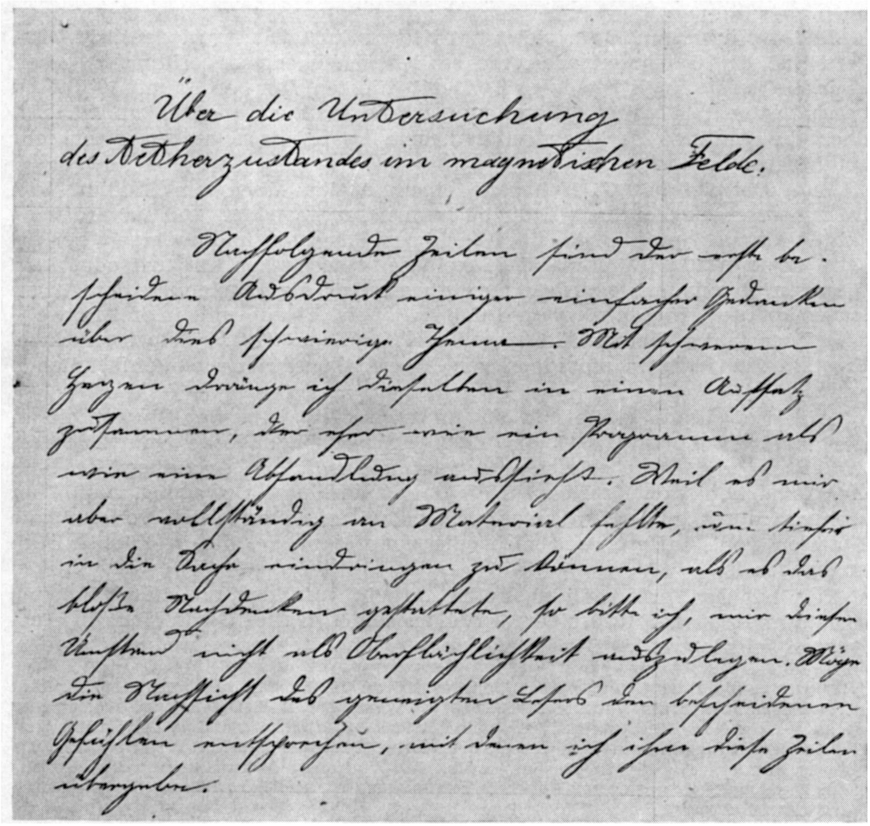
1
Albert Einstein's 'First' Paper*

In 1894 or 1895, the young Albert Einstein wrote an essay on 'The Investigation of the State of Aether in Magnetic Fields.' He sent the essay, most probably his 'first' scientific work, with a letter to his uncle Cäsar Koch. Both items are presented in this article with some comments on the origins of Einstein's ideas on Special Relativity.

Albert Einstein always maintained that the trend of thinking that ultimately led to his work '*Zur Elektrodynamik bewegter Körper*' ('On the Electrodynamics of Moving Bodies')[1] had already begun when he was an adolescent young man. In conversations and interviews at various times, several people sought to find out from Einstein himself about his intellectual and scientific development in order to fix the chronology of the conception, gestation and birth of the Special Theory of Relativity. We know very little about Einstein as a boy and young scholar other than what he has himself mentioned in scattered writings or told his biographers and interviewers.

Gerald Holton, in his article 'Influences on Einstein's Early Work in Relativity Theory,' reported on his search in documents, diaries, notebooks, correspondence, and unpublished manuscripts in the Einstein archives at Princeton and other source materials for any indications relating to Einstein's 1905 paper on relativity,

*During the summer semester in June 1970, I gave a series of lectures at the International Solvay Institutes of the *Université libre de Bruxelles* on the historical development of the quantum and relativity theories. One of my auditors was a young man, Jean Ferrard, whom Professor Jean Pelseneer introduced me to as the grandson of Madame Suzanne Koch-Gottschalk, the daughter of Cäsar Koch and thus Einstein's cousin. Monsieur Ferrard arranged my meeting with Madame Suzanne Koch-Gottschalk, during which she told me that she had a box of papers in which there might be some Einstein documents and if I would help her in sorting them out. I was very excited by this opportunity, and went through the papers in the box; contained in it were Einstein's essay, discussed here, and the covering letter to his uncle. I told Madame Suzanne Gottschalk about the importance of these documents, and asked her permission to publish them, which she readily granted. I wrote this article and made copies of the Einstein documents I had found in the Gottschalk family box and personally gave them to Miss Helen Dukas in Princeton in May 1970; she and Otto Nathan, executor of the Einstein Estate, gave me permission to publish my article. After completing this essay, I sent a preprint of it to Freeman Dyson (as I did of all my papers for his comments); he replied to me at once, and said among other things: 'This paper is like the discovery of Linear B by Michael Ventris, and shows how humble are the origins of modern science. It is an important find; publish it immediately! Freeman.' It was published in *Physikalische Blätter* **27**, 385 (1971) and as Report No. CPT-82; AEC-31, January 8, 1971, of the Center for Particle Theory, The University of Texas at Austin. I have included this essay in this volume because of its historical interest.


Über die Untersuchung
des Aetherzustandes im magnetischen Felde.

The introduction to Einstein's essay on 'The Investigation of the State of Aether in Magnetic Fields,' which he wrote at the age of 15 or 16 and sent to his uncle Cäsar Koch with a covering letter.

concluding that 'there is no contemporaneous draft or manuscript from which one might learn something of the genesis of the paper.'[2] Holton surveyed various books and authors that probably might have influenced the young Einstein ever since he came to the Aargau Cantonal School in Aarau in 1895,[3] and looked for some remarks or evidence that might have started Einstein's thinking on relativity. In these and the two notebooks of lecture notes which Einstein kept at Zurich during the period 1897 to 1900, Holton drew a blank. After a thorough and intensive search Holton decided that August Föppl's successful book '*Einführung in die Maxwellsche Theorie der Elektrizität*' ('Introduction to Maxwell's Theory of Electricity'), first published in 1894,[4] had a decisive influence on Einstein, that Föppl was the 'almost forgotten teacher.' Föppl, as Holton noted, was called in 1894 to the Technische Hochschule at Munich, where the young Einstein was then living. The suggestion is enticing that Einstein became familiar with Föppl's book shortly after it was published, learned from it the fundamentals of the electromagnetic theory of Maxwell and Hertz, was influenced by it in his formative years, and was probably inspired by it to work on the theory of relativity. The headings like 'The Electrodynamics of Moving Conductors,' 'Electromagnetic Force Induced by Movement,' and 'Relative and Absolute Motion in Space' which occur in Föppl's book are indeed suggestive of an influence on Einstein. We should, however, remember that these topics were frequent enough in the scientific literature of the 1890's and the early 1900's, as the electrodynamics of moving bodies was one of the central problems of physics. Notation is often indicative of influence, and Einstein used Hertz' notation in his 1905 paper on relativity, not the Heaviside notation used by Föppl, indicating that Hertz' direct influence on Einstein might have been greater than it is generally assumed.[4a]

I have recently found a 'paper' (an essay) which the young Albert Einstein wrote in 1894 or 1895.[5] He sent the essay entitled '*Über die Untersuchung des Aetherzustandes im magnetischen Felde*' ('Concerning the Investigation of the State of Aether in Magnetic Fields') with a covering letter to his maternal uncle, Cäsar Koch, who was living in Antwerp, Belgium, at that time.[6*]

The letter to Cäsar Koch and the essay were assigned the date of '1894 or 1895' by Einstein himself in 1950.[7] It is quite evident from the letter that he wrote the essay (during the year he spent in Milan) before he went to Zurich for the entrance examination of the 'Polytechnic' [E.T.H., *Eidgenössiche Technische Hochschule*, Zurich — the Swiss Federal Institute of Technology].[8] The essay therefore predates all source materials to which references have been made in connection with the origins of the theory of relativity. Before presenting Einstein's letter and essay, let us review some of the autobiographical and interview comments which are on record; all of them dwell upon Einstein's recollection of the vague beginnings of the 'relativity problem' going back to his Aarau days.

Thus the psychologist Max Wertheimer recalled 'those wonderful days,' beginning in 1916, when he questioned Einstein for hours on end alone in his study, and heard from him the story of the 'dramatic developments which culminated in the theory of relativity.' Wertheimer probed Einstein for the 'concrete events in his thought,' and Einstein described to him the genesis of each equation in great and specific detail.[9]

By the time the conversations with Max Wertheimer took place, Einstein had already completed the main edifice of his theory of gravitation and the general theory of relativity,[10] and he could take a long look back and reminisce about a glorious intellectual adventure. As Wertheimer recalls, 'The problem started when Einstein was sixteen years old, a pupil in the Gymnasium (Aarau, Kantonschule). He was not an especially good student, unless he did productive work on his own account. This he did in physics and mathematics, and consequently he knew more about those subjects than his classmates. It was then that the great problem really started to trouble him. He was intensely concerned with it for seven years; from the

---

*I presented photocopies of these documents to Miss Helen Dukas in May 1970 for the Einstein Archives in Princeton. I am grateful to Miss Dukas and the Executor of the Estate of Albert Einstein, Dr. Otto Nathan, for permission to publish these items.

moment however that he came to question the customary concept of time, it took him only five weeks to write his paper on relativity....'[11]

On 4 February 1950, in the first of several visits that he made to Einstein in Princeton during the period 1950–1954, R.S. Shankland asked Einstein how long he had worked on the Special Theory of Relativity before 1905. Einstein told him that he had started on the problem at the age of 16, already as a student when he could devote only part of his time to it, and worked on it for ten years. He made many fruitless attempts to develop a theory consistent with the experimental facts, but they had to be abandoned, 'until it came to me that time was suspect!'[12]

Einstein, in his conversation with Shankland, commented at length on the nature of mental processes, and emphasized that our minds do not seem to move step by step to the solution of a problem; rather, they take a devious route. 'It is only at the last that order seems at all possible in a problem,' said Eintein.[13] Of a later interview on 24 October 1952, Shankland reports, 'I asked Professor Einstein about the three famous 1905 papers [*Annalen der Physik*, **17**, 132, 549, 891 (1905)] and how they all appeared to come at once. He told me that the work on special relativity "had been his life for over seven years and that this was the main thing." However, he quickly added that the photoelectric effect paper was also the result of five years pondering and attempts to explain Planck's quantum in more specific terms. He gave me the distinct impression that the work on Brownian motion was a much easier job. "A simple way to explain this came to me, and I sent it off." '[14]

So again it was relativity, the problem of the electrodynamics of moving bodies, that went farthest back in his memory. Not only did Einstein have curiosity about the workings of nature, he had also acquired some knowledge of the essentials of physics and mathematics quite early in school. His remarks indicate that, even as a boy of sixteen, he had recognized the intellectual challenge of some fundamental problems of physics.[15]

Excitement about natural phenomena had come to Einstein early. At the age of 4 or 5, he had received a compass from his father to play with. The sense of wonder, of a 'secret power behind the movement of the needle,' which he experienced as a child remained a deep and lasting memory with him.[16] The various business crises of his father, which affected the fortunes of the family, did not destroy the atmosphere of free thought, experience, and sense of mystery about nature in which Einstein grew up. In 1889, at the age of 10 Albert Einstein entered the *Luitpold Gymnasium* in Munich. His work at the *Gymnasium* was a mechanical routine; but still, at the age of 12, he experienced the excitement and beauty of geometry when he came across an old textbook on Euclidean plane geometry at the school.

Of his boyhood studies, Einstein recalled in his autobiographical notes: 'At the age of 12–16 I familiarized myself with the elements of mathematics together with the principles of differential and integral calculus. In doing so I had the good fortune of hitting up books which were not too particular in their logical rigour, but which made up for this by permitting the main thoughts to stand out clearly and synoptically. This occupation was, on the whole, truly fascinating; climaxes were reached whose impression could easily compete with that of elementary geometry — the basic idea of analytical geometry, the infinite series, the concepts of differential and integral. I also had the good fortune to know the essential results and methods of the entire field of the natural sciences in an excellent popular exposition, which limited itself almost throughout to qualitative aspets ([Aaron] Bernstein's *People's Books on Natural Science*, a work of 5 or 6 volumes), a work which I read with breathless attention. I had also already studied some theoretical physics when, at the age of 17, I entered the Polytechnic Institute of Zürich as a student of mathematics and physics.'[17] Einstein also recalled that 'at the age of 13 I read with enthusiasm Ludwig Büchner's *Force and Matter*, a book which I later found to be rather childish in its ingenuous realism.'[18]

On account of business difficulties his father left Munich in 1894 for Milan, but Einstein stayed on in a pension to complete his studies at school. He found the mechanical routine of his academic life at the *Gymnasium* intolerable, and a few months later he joined his parents in Milan. He had left the unpleasant rigors and discipline of the German *gymnasium*, but had also left the school in Munich without a diploma. Einstein was fifteen years old.

Einstein spent a year with his parents in Milan, and during this time thought about pursuing higher education in theoretical physics. Having no diploma from the *Gymnasium*, he thought of gaining admission to the Swiss Federal Institute of Technology in Zurich by taking the entrance examination. Later on, he recalled:

'As a sixteen-year-old I came to Zurich from Italy in 1895, after I had spent one year without school and teachers in Milan with my parents. My aim was to gain admission to the Polytechnic, but it was not clear to me how I should attain this, I was a self-willed but modest young man, who had obtained his fragmentary knowledge of the relevant fundamentals [mainly] by self-study. Avid for deeper understanding, but not very gifted in being receptive, studies did not appear to me to be an easy task. I appeared for the entrance examination of the engineering department with a deep-seated feeling of insecurity. Even though the examiners were patient and understanding, the examination painfully revealed to me the gaps in my earlier training. I thought it was only right that I failed. It was a comfort, however, that the physicist H.F. Weber informed me that I could attend his lectures if I stayed in Zurich. The director, Professor Albin Herzog, however, recommended me to the Cantonal School in Aarau, from where after one year's study I was graduated. On account of its liberal spirit and genuine sincerity, and teachers who did not lean on external authority of any kind, this school has left on me an unforgettable impression. Compared to the six years of schooling in an authoritatively run German *gymnasium* I became intensely aware of how much education leading to independent activity and individual responsibility is to be preferred to the education which relies on drill, external authority, and ambition. Real democracy is not an empty illusion.

'*During this year in Aarau came to me the question: If one follows a light beam with the speed of light, then one would obtain a time-indepedent wave field. However, such a thing does not exist! This was the first childish thought-experiment which had* *something to do with the Special Theory of Relativity.* Invention is not the result of logical thinking, even though the final result has to be formulated in a logical manner.'[19] (My italics.)

Einstein tried to imagine what he would observe if he were to travel through space with the same velocity as a beam of light. According to the usual idea of relative motion, it would seem that the beam of light would then appear as a spatially oscillating static electromagnetic field. But such a concept was unknown to physics and at variance with Maxwell's theory. Einstein began to suspect that the laws of physics, including those concerning the propagation of light, must remain the same for all observers however fast they move relative to one another.[20]

When Wertheimer asked Einstein if already at that time he had some idea of the invariance of the velocity of light for all observers in uniform motion, Einstein replied, 'No, it was just a curiosity. That the velocity of light could differ depending on the movement of the observer was somehow characterized by doubt. Later developments increased that doubt.'[21]

Says Wertheimer: 'Light did not seem to answer when one put such questions. Also light, just as mechanical processes, seemed to know nothing of a state of absolute movement or of absolute rest. This was interesting, exciting.

'Light was to Einstein something very fundamental. At the time of his studies at the *Gymnasium* [Aarau], the aether was no longer being thought of as something mechanical, but as "the mere carrier of electrical phenomena." '

Einstein's essay on the state of the aether in magnetic fields, presented in the following, refers to his familiarity with the experiments, and deals rather vaguely with the connection between the aether and electromagnetic phenomena. In his essay, presented here, Einstein proposed a method for detecting elastic deformations of the aether by sending light rays into the vicinity of the current-carrying wire. In his essay, Einstein raised the following main questions: (i) How does a magnetic field, which is generated when a current is turned on, affect the surrounding aether? (ii) How does this magnetic field, in turn, affect the current itself? Einstein believed in the existence of an aether at that time, and regarded it as an elastic medium; he wondered in particular how 'the three components of elasticity act on the velocity of the aether wave' which is generated when the currrent is turned on. His main conclusion was that 'Above all, it ought to be [experimentally] shown that there exists a passive resistance to the electric current's ability for generating a magnetic field; [this resistance] is proportional to the length of the wire and independent of the cross section and the material of the conductor.' Thus the young Einstein independently discovered the qualitative properties of self-induction, and it seems clear that Einstein was not yet familiar with the earlier work on this phenomenon, though at that time he knew that light is an electromagnetic phenomenon but was not yet familiar with Maxwell's theory.[21a]

The problems he thought about at Aarau clearly occurred to him after writing this essay. It is quite possible that sometime during his stay in Munich, Milan, Aarau or Zurich (that is, during the period 1894–1900), or even perhaps in Berne during 1900–1905, Föppl's book[4] (to which Holton[2] attaches great importance as a possible influence on Einstein's early work on relativity theory) fell into Einstein's hands. It is important to emphasize, however, that Einstein does mention 'the wonderful experiments of Hertz' in his essay, and he continued to mention Hertz among his 'unforgotten' teachers like Helmholtz, Maxwell, Boltzmann, and Lorentz.[22] Whatever the real influences might have been on the genesis of Einstein's 1905 paper,[23] the following bits of "Einsteiniana' are the earliest available record of Einstein's intellectual adventure from his own hand, and they have therefore some historical interest.

Letter to Cäsar Koch*

1894 or 1895. A. Einstein. (Date recalled in 1950.)

My dear Uncle:

I am really very happy that you are still interested in the little things I am doing and working on, even though we could not see each other for a long time and I am such a terribly lazy correspondent. I always hesitated to send you this [attached] note; because it deals with a very special topic, and besides it is still rather naive and imperfect, as is to be expected from a young fellow like myself. I shall not be offended at all if you don't read the stuff; but you must recognize it at least as a modest atttempt to overcome the laziness in writing which I have inherited from both of my dear parents....

As you probably already know, I am now expected to go to the Polytechnic in Zurich. However, it presents serious difficulties because I ought to be at least two years older for that. We shall write to you in the next letter what happens in this matter.

Warm greetings to dear aunt and your lovely children,

from your
Albert

*Einstein's maternal uncle; sometimes in letters and addresses the name has been spelled with the French accent as 'César.' (My translation of the letter.)

Concerning the Investigation of the State of Aether in Magnetic Fields*

The following lines are the first modest expression of some simple thoughts on this difficult subject. With much hesitation I am compressing them into an essay which looks more like a program than a paper. Since I completely lacked the materials to penetrate the subject more deeply than was permitted by reflection alone, I ask that this circumstance should not be ascribed to me as superficiality. I hope the indulgence of the interested reader will correspond to the humble feelings with which I offer him these lines.

When the electric current comes into being, it immediately sets the surrounding aether in some kind of instantaneous motion, the nature of which has still not been exactly determined. In spite of the continuation of the cause of this motion, namely the electric current, the motion ceases, but the aether remains in a potential state and produces a magnetic field. That the magnetic field is a potential state [of the aether] is shown by the [existence of a] permanent magnet, since the principle of conservation of energy excludes the possibility of a state of motion in this case. The motion of the aether, which is caused by an electric current, will continue until the acting [electro-] motive forces are compensated by the equivalent passive forces which arise from the deformation caused by the motion of the aether itself.

The marvellous experiments of Hertz have most ingeniously illuminated the dynamic nature of these phenomena — the propagation in space, as well as the qualitative identity of these motions with light and heat. I believe that for the understanding of electromagnetic phenomena it is important also to undertake a comprehensive experimental investigation of the potential states of the aether in magnetic fields of all kinds — or, in other words, to measure the elastic deformations and the acting deforming forces.

Every elastic change of the aether at any (free) point in a given direction should be determinable from the change which the velocity of an aether wave undergoes at this point in that direction. The velocity of a wave is proportional to the square root of the elastic forces which cause [its] propagation, and inversely proportional to the mass of the aether moved by these forces. However, since the changes of density caused by the elastic deformations are generally insignificant, they may probably be neglected in this case also. It could therefore be said with good approximation: The square root of the ratio of the change of velocity of propagation (wavelength) is equal to the ratio of the change of the elastic force.

I dare not decide as to which type of aether waves, whether light or electrodynamic, and which method of measuring the wavelength is most appropriate for studying the magnetic field; in principle, after all, this makes no difference.

If a change of wavelength in the magnetic field can be detected at all in any given direction, then the question can be experimentally decided whether only the component of the elastic state in the direction of the propagation of the wave influences

*My translation of Einstein's essay.

the velocity of propagation, or the components perpendicular to it also do; since it is known *a priori* that in a uniform magnetic field, whether it is cylindrical or pyramidal in form, the elastic states at a point perpendicular to the direction of the lines of force are completely homogeneous, but different in the direction of the lines of force. Therefore if one lets waves propagate that are polarized perpendicularly to the direction of the lines of force, then the direction of the plane of oscillation would be important for the velocity of propagation — that is if the component of the elastic force perpendicular to the propagation of a wave at all influences the velocity of propagation. However, this probably might not be the case, although the phenomenon of double diffraction seems to indicate this.

Thus after the question has been answered as to how the three components of elasticity affect the velocity of an aether wave, one can proceed to the study of the magnetic field. In order to understand properly the state of the aether in it [the magnetic field], three cases ought to be distinguished:

1. The lines of force come together at the North pole in the shape of a pyramid.
2. The lines of force come together at the South pole in the shape of a pyramid.
3. The lines of force are parallel.

In these cases the velocity of propagation of a wave in the direction of the lines of force and perpendicular to them has to be examined. There is no doubt that the elastic deformations as well as the cause of their origin will be determined [by these experiments], provided sufficiently accurate instruments to measure the wavelength can be constructed.

The most interesting, but also the most difficult, task would be the direct experimental study of the magnetic field which arises around an electric current, because the investigation of the elastic state of the aether in this case would allow us to obtain a glimpse of the mysterious nature of the electric current. This analogy also permits us to draw definite conclusions concerning the state of the aether in the magnetic field which surrounds the electric current, provided of course the experiments mentioned above yield any result.

I believe that the quantitative researches on the absolute magnitudes of the density and the elastic force of the aether can only begin if qualitative results exist that are connected with established ideas. Let me add one more thing. If the wavelength does not turn out to be proportional to $\sqrt{A+k}$ [*sic*], then the reason (for that) has to be looked for in the change of density of the moving aether caused by the elastic deformations; here A is the elastic aether force, *a priori* a constant which we have to determine empirically, and k the (variable) strength of the magnetic field which, of course, is proportional to the elastic forces in question that are produced.

Above all it must be demonstrated that there exists a passive resistance to the electric current for the production of the magnetic field, that is proportional to the length of the path of the current and independent of the cross section and the material of the conductor.

Notes and References

1. A. Einstein, *Annalen der Physik*, Ser. 4, **17**, pp. 891–921, 1905.
2. Gerald Holton, 'Influences of Einstein's Early Work in Relativity Theory,' *The American Scholar*, **37**, No. 1, pp. 59–79, Winter, 1967–68.
3. Albert Einstein was a pupil in the third and fourth classes of the Aargau Cantonal School in Aarau from October 1895 to early fall of 1896. In October 1896 Einstein enrolled at the E.T.H., Zurich, to study for a *Fachlehrer* (specialist teacher) diploma in mathematical physics, and was graduated in August 1900. (See Carl Seelig, *Albert Einstein*, Staples Press, London, 1956.) In his article (*ibid.*, p. 63) Holton remarks: 'As Besso wrote (in his notes of August 1946 for Strickelberg's article on Einstein in Switzerland), Einstein came to the Aarau Kanton-School in 1896. . . .' There is a slight confusion of dates in this. Both Seelig and Einstein are correct about the dates.
4. August Föppl, *Einführung in die Maxwellsche Theorie der Elektrizität*, Druck und Verlag von B.G. Teubner, 1894.

4a. Heinrich Hertz' *Untersuchungen über die Elektrischen Kraft* was published in 1892; the first English edition of his *Electric Waves* was published in 1894 (McMillan and Co. Ltd.). Hertz died in 1894. Paul Drude's book *Physik des Aethers* was also published in 1894 (Verlag von Ferdinand Enke, Stuttgart, 1894).

Einstein's essay, presented in this article, clearly indicates that his interest in electromagnetism was aroused by the 'marvellous' experiments of Heinrich Hertz. These experiments, since Faraday's early work, were the most important in the field of electromagnetism and were justly so celebrated at the time. Faraday had discovered the law of electromagnetic induction in 1834, and it was this law that guided Einstein in his work on Special Relativity. Einstein built his theory on experimental facts. He starts his 1905 paper by pointing out that the law of induction contains an asymmetry which is artificial, and does not correspond to facts. Empirical observation shows that the current induced depends only on the relative motion of the conducting wire and the magnet, while the usual theory explains the effect in quite different terms according to whether the wire is at rest and the magnet moving or vice versa. At the time of Einstein's writing the law of induction was about 70 years old, and 'everybody had known all along that the effect depended on relative motion, but nobody had taken offence at the theory not accounting for this circumstance.' (See Max Born, Physics and Relativity, in *Physics in My Generation*, Springer Verlag New York, 1969.)

5. I am grateful to Madame Suzanne Koch-Gottschalk and Jean Ferrard for allowing me to examine the letters and papers relating to Einstein in the possession of their family. Madame Gottschalk has very kindly allowed me to publish the translation of Einstein's essay and to report on my findings for scientific purposes.
6. Cäsar and Jakob Koch were the two brothers of Einstein's mother Pauline. Jakob lived in Zurich and his name occurs several times in the Einstein-Besso correspondence. Cäsar Koch seems to have been Einstein's favorite relative. In one of the letters to Cäsar, Einstein remarks: '. . . . *Bist Du mir doch immer der Liebste in der Familie gewesen.*' After his marriage to Mathilde Lévy at Basle in 1888, Cäsar went to Buenos Aires; following a sojourn there and return to Basle, he settled in Antwerp, Belgium, around 1891, and moved to Brussels after the First World War. He was a merchant of commodities. Cäsar Koch was very fond of Albert Einstein and encouraged him in his boyhood studies. Einstein visited the Koch family when he attended the first Solvay conference in Brussels in 1911. In the 1920's when Einstein used to visit Ehrenfest in Leiden and gave lectures there, or en route to Paris from Berlin, he always visited his uncle Cäsar. These personal contacts were continued when Einstein attended the fifth and sixth Solvay conferences in Brussels, in 1927 and 1930 respectively, and during

the months he spent as a refugee from Germany at Le Coq sur Mer near Ostende before his final departure from Europe to the United States. Later on, affectionate correspondence between Einstein, his wife Elsa, his sister Maja, and the Koch family was maintained.

I am grateful to Madame Suzanne Gottschalk, daughter of Cäsar Koch, for conversations about Einstein and her family, and for showing me numerous letters and photographs.

7. I have not been able to discover the identity of the person who showed Einstein these documents in 1950. Neither the owners of the documents nor Miss Helen Dukas, Einstein's former secretary, have any recollection of who this person was, nor could they offer any reasonable guess about his identity.
8. The choice of the E.T.H. in Zurich for Einstein's higher studies was made by his father Hermann and uncle Jakob Einstein. The two brothers at one time had founded a small engineering factory for making dynamos, measuring instruments and arc lamps, and Einstein's initial plan was to study engineering in Zurich.
9. Max Wertheimer, *Productive Thinking*, edited by Michael Wertheimer, Enlarged Edition 1959, Harper & Row Publishers, New York and Evanston, p. 213.
10. In 1905 Einstein continued the theme of his relativity paper by discussing the dependence of the inertia of a body on its energy (*Annalen der Physik*, ser. 4, vol. 18, pp. 639–641). In 1907 Einstein wrote on the possibility of a new test of the principle of relativity (*Annalen der Physik*, ser. 4, vol. 23, pp. 197–198), the inertia of energy as a consequence of the relativity principle (*Annalen der Physik*, ser. 4, vol. 23, pp. 371–384), and '*Relativitätsprinzip und die aus demselben gezogenen Folgerungen*' (*Jahrbuch der Radioaktivität,* vol. 4, pp. 411–462, and vol. 5, pp. 98–99). In the last paper he explicitly stated the equivalence of inertial and gravitational mass, and gave the famous equation for mass in terms of energy. Einstein returned to the ideas of this paper in 1911 when he wrote on the influence of gravity on light (*Annalen der Physik*, ser. 4, vol. 35, pp. 898–908). The theme of relativity and gravitation was taken up in 1912, with papers on the velocity of light in a gravitational field (*Annalen der Physik*, ser. 4, vol. 38, pp. 355–369), the theory of a static gravitational field (*Annalen der Physik*, ser. 4, vol. 38, pp. 443–458), and replies to remarks of M. Abraham in short notes (*Annalen der Physik*, ser. 4, vol. 8, pp. 1059–1064; vol. 39, p. 704). A major summing up of the ideas expressed in these papers and approaches to the general theory of relativity and gravitation were made with Marcel Grossmann in '*Entwurf einer Verallgemeinerten Relativitätstheorie und eine Theorie der Gravitation*' (*Zeitschrift für Mathematik and Physik*, vol. 62, pp. 225–261). Einstein presented a lecture on the physical foundations of the new theory of gravitation in an address to the *Naturforschende Gesellschaft*, Zurich, 9 September 1913 (*Vierteljahrsschrift*, vol. 58, pp. 284–290), and continued the theme two weeks later in a lecture at the 85th *Versammlung Deutscher Naturforscher* in Vienna on 21 September 1913. In 1914, Einstein wrote on the formal foundations of general relativity theory (*Sitzungsberichte der Preussischen Akademie der Wissenschaften*, part 2, pp. 1030–1085, 1914), gave several lectures on the problem of gravitation and relativity, and published a paper with M. Grossmann on the general convariance properties of the field equations of the theory of gravitation (*Zeitschrift für Mathematik und Physik*, vol. 63, pp. 215–225). Einstein continued to write on general relativity during the year 1915, and published new ideas on the application of the theory of astronomy; he also explained the perihelion motion of mercury on the basis of the general theory. Then in 1916 his great paper on the complete general theory of relativity was published: '*Grundlage der allgemeinen Relativitäts-theorie*' (*Annalen der Physik*, ser. 4, vol. 49, pp. 769–822). In an important sense this was the culmination

of the intellectual adventure on which Einstein had started since the time he wrote to his uncle Cäsar Koch in 1894 or 1895.

11. Max Wertheimer, *ibid.*, p. 214.
12. R.S. Shankland, 'Conversations with Albert Einstein,' *American Journal of Physics*, vol. 31, pp. 47–57, 1963, p. 48.
13. R.S. Shankland, *ibid.*, p. 48.
14. R.S. Shankland, *ibid.*, p. 56.
15. Apart from questions concerning light and the electrodynamics of moving bodies, Einstein went to Aarau 'with the [then much debated] questions concerning the palpability [*Greifbarkeit*] of ether and of atoms' in mind. (For the quotation from Besso in this remark, see Holton, *ibid.*, p. 63.)
16. Autobiographical notes by Albert Einstein in *Albert Einstein, Philosopher–Scientist*, edited by Paul Arthur Schilpp (originally in Library of Living Philosophers, 1949), Harper Torchbooks Science Library, New York, 1959, p. 9.
17. Autobiographical notes by Albert Einstein, *ibid.*, p. 15.
18. Carl Seelig, *Albert Einstein*, Staples Press, London, 1956, p. 12.
19. Albert Einstein, *Autobiographische Skizze*; perhaps one of the last writings of Einstein (written in March 1955), was published in Fall 1955 in '*Schweizerische Hochschulzeitung*,' *Festnummer* 1855–1955, on the occasion of the centennial jubilee of the E.T.H. in Zurich. In this autobiographical sketch, Einstein recalled some touching memories of his life in Switzerland. This sketch was included in *Helle Zeit — Dunkle Zeit, In Memoriam Albert Einstein*, edited by Carl Seelig, Europa Verlag, Zurich, 1956, pp. 9–17. See pp. 9–10 for the quotation.
20. *Einstein: The Man and His Achievement*, a series of broadcasts on the BBC Third Programme, edited by G.J. Whitrow, BBC, London, 1967.
21. Max Wertheimer, *ibid.*, p. 215.

21a. A. Pais, *Subtle is the Lord*, Oxford University Press, 1982, p. 131.

22. Louis Kollross, '*Albert Einstein en Suisse—Souvenirs*,' in '*Funfzig Jahre Relativitätstheorie*' (Bern, 11–16 July 1955), *Helvetica Physica Acta, Supplementum IV*, 1956, see pp. 274–275; also published as '*Erinnerungen eines Kommilitonen*' in '*Helle Zeit — Dunkle Zeit*,' *ibid.*, see p. 22.
23. In an undated letter, probably sometime after 6 March 1905, Einstein wrote to his friend Conrad Habicht in Schiers; 'But why have you not yet sent me your thesis? Don't you know, you wretch, that I should be one of the few fellows who would read it with interest and pleasure? I can promise you in return four works, the first of which I shall soon be able to send you as I am getting some free copies. It deals with the radiation and energy characteristics of light and is very revolutionary, as you will see if you send me your work in advance. The second study is a determination of the true atomic dimensions from the diffusion and inner friction of diluted liquid solutions of neutral matter. The third proves that on the premise of the molecular theory of induction, particles of the size 1/1000 mm., when suspended in liquid, must execute a perceptible irregular movement which is generated by the movement of heat. Movements of small, lifeless, suspended particles have in fact been examined by physiologists and these movements have been called by them "the Brownian movement." *The fourth study is still a mere concept: the electrodynamics of moving bodies by the use of a modification of the theory of space and time. The purely cinematic part of this work will undoubtedly interest you.*' (My italics.)

    The fourth study, to which Einstein refers, was his paper on the Special Theory of Relativity. It was completed in Berne in June 1905, and received by the editor of *Annalen der Physik* on 30 June 1905. It is indeed quite remarkable that even at this late date (sometime after 6 March 1905), Einstein refers to his study as 'still a mere concept.' This concept, however, had now been growing within him for almost ten years. On March 11, 1952, Albert Einstein wrote to Carl Seelig: 'Between the conception of the idea of this special relativity theory and the completion of the corresponding publication, there elapsed five or six weeks. But [he added rather cryptically] it would be hardly correct to consider this as a birthdate, because earlier the arguments and building blocks were being prepared over a period of years, although without bringing about the fundamental decision.' (See Ref. 2, p. 60.) Michele Besso, Einstein's friend and colleague at the Patent Office in Berne, was party to the 'fundamental decision,' the final progress of Einstein's conception, and its publication. In concluding his paper, Einstein wrote, 'I wish to say that in working at the problem dealt with here, I have had the loyal assistance of my friend and colleague M. Besso, and I am indebted to him for several valuable suggestions.'

1894 oder 95. A. Einstein. (Datum 1950 nachgeholt.)

Mein lieber Onkel!

Es freut mich wirklich sehr, dass Du Dich für mein bischen Thun und Treiben noch interessierst, trotzdem wir uns so lange nicht sehen durften und ich so grässlich fauler Briefschreiber bin. Und doch zögerte ich immer, Dir dieses Schreiben hier zu schicken. Denn es behandelt ein ein [sic] sehr speziales Thema, und ist ausserdem, wie es sich für so einen jungen Kerl wie mich von selbst versteht, noch ziemlich naiv und unvollkommen. Wenn Du das Zeug gar nicht liest, nehme ich Dirs durchaus nicht übel; Du musst es aber doch zum mindesten als einen schüchternen Versuch anerkennen, die von meinen beiden lieben Eltern geerbte Schreibfaulheit zu bekämfen - - -

Wie Du schon wissen wirst soll ich jetzt auf das Polytechnikum nach Zürich kommen. Die Sache stösst aber auf bedeutende Schwierigkeiten, da ich dazu eigentlich zwei Jahre mindestens älter sein sollte. Im nächsten Brief schreiben wir Dir, was aus der Sache wird.

Innige Grüsse der lieben Tante und Deinen herzigen Kinderchen

von Deinem
Albert

Über die Untersuchung des Aetherzustandes im magnetischen Felde

Nachfolgende Zeilen sind der erste bescheidene Ausdruck einiger einfacher Gedanken über dies schwierige Thema. Mit schwerem Herzen dränge ich dieselben in einen Aufsatz zusammen, der eher wie ein Programm als wie eine Abhandlung aussieht. Weil es mir aber vollständig an Material fehlte, um tiefer in die Sache eindringen zu können, als es das blosse Nachdenken gestattete, so bitte ich, mir diesen Umstand nicht als Oberflächlichkeit auszulegen. Möge die Nachsicht des geneigten Lesers den bescheidnen Gefühlen entsprechen, mit denen ich ihm diese Zeilen übergebe.

Der elektrische Strom setzt bei seinem Entstehen den umliegenden Äther in irgend eine, bisher ihrem Wesen nach noch nicht sicher bestimmte, momentane Bewegung. Trotz Fortdauer der Ursache dieser Bewegung, nämlich des elektrischen Stroms, hört die Bewegung auf, der Äther verbleibt in einem potentiellen Zustande und bildet ein magnetisches Feld. Dass das magnetische Feld ein potentieller Zustand sei, beweisst der permanente Magnet, da das Gesetz von der Erhaltung der Energie hier die Möglichkeit eines Bewegungszustandes ausschliesst. Die Bewegung des Äthers, welche durch einen elektrischen Strom bewirkt wird, wird so lange dauern, bis die wirkenden motorischen Kräfte durch äquivalente passive Kräfte kompensiert werden, welche von der durch die Bewegung des Äthers selbst erzeugten Deformationen herrühren.

Die wunderbaren Versuche von Hertz haben die dynamische Natur dieser Erscheinungen, die Fortpflanzung im Raume, sowie die qualitative Identität dieser Bewegungen mit Licht und Wärme aufs genialste beleuchtet. Ich glaube nun, dass es für die Erkenntnis der elektromagnetischen Erscheinungen von Wichtigkeit wäre, auch die potentiellen Zustände des Äthers in magnetischen Feldern aller Art einer umfassenden experimentellen Betrachtung zu unterziehen, oder mit anderen Worten, die elastischen Deformationen und die wirkenden deformierenden Kräfte zu messen.

Jede elastische Veränderung des Äthers an irgend einem (freien) Punkte in einer Richtung muss sich konstatieren lassen aus der Veränderung, welche die Geschwindigkeit einer Ätherwelle an diesem Punkte in dieser Richtung erleidet. Die Geschwindigkeit einer Welle ist proportional der Quadratwurzel der elastischen Kräfte, welche zur Fortpflanzung dienen, und umgekehrt proportional der von diesen Kräften zu bewegenden Äthermassen. Da jedoch die durch die elastischen Deformationen hervorgerufenen Veränderungen der Dichte meist nur unbedeutend sind, so wird man sie auch in diesem Falle wahrscheinlich vernachlässigen dürfen. Man wird also mit grosser Annäherung sagen können: Die Quadratwurzel aus dem Verhältnis der Veränderung der Fortpflanzungsgeschwindigkeit (Wellenlänge) ist gleich dem Verhältnis der Veränderung der elastischen Kraft.

Was für eine Art von Ätherwellen, ob Licht oder elektrodynamische, und was für eine Methode der Messung der Wellenläange für die Untersuchung des magnetischen Feldes am geeignetsten sei, wage ich nicht zu entscheiden; im Prinzip ist es ja schliesslich gleich.

Zunächst kann, wenn überhaupt eine Veränderung der Wellenlänge im magnetischen Feld in irgendeiner Richtung sich konstatieren lässt, experimentell die Frage gelöst werden, ob nur die Komponente des elastischen Zustandes in der Richtung der Fortpflanzung der Welle oder auch die dazu senkrechten Komponenten eine Wirkung auf die Fortpflanzungsgeschwindigkeit ausüben, da a priori klar ist, dass in einem regelmässigen magnetischen Feld, sei es zylinder- oder pyramidenförmig, die elastischen Zustände an einem Punkt senkrecht zur Richtung der Kraftlinien vollständig homogen sind und anders in der Richtung der Kraftlinien. Lässt man daher senkrecht zur Richtung der Kraftlinien polarisierte Wellen durchdringen, so wäre für die Fortpflanzungsgeschwindigkeit die Richtung der Schwingungsebene von Bedeutung — wenn die zur Fortpflanzung einer Welle senkrechte Komponente der elastischen Kraft wirklich auf die Geschwindigkeit der Fortpflanzung einen Einfluss ausübt. Dies dürfte jedoch wahrscheinlich nicht der Fall sein, trotzdem das Phänomen der Doppelbrechung darauf hinzuweisen scheint.

Nachdem so die Frage entschieden wäre, wie die drei Komponenten der Elastizität auf die Geschwindigkeit einer Ätherwelle einwirken, kann zur Untersuchung des magnetischen Feldes geschritten werden. Um den Zustand des Äthers in demselben recht begreifen zu können dürften drei Fälle unterschieden werden:

1. Kraftlinien, die sich pyramidenartig am Nordpol vereinigen.
2. Kraftlinien, die sich pyramidenartig am Südpol vereinigen.
3. Parallele Kraftlinien.

In diesen Fällen ist die Fortpflanzungsgeschwindigkeit einer Welle in der Richtung der Kraftlinien und senkrecht dazu zu untersuchen. Unzweifelhaft müssen sich so die elastischen Deformationen samt ihrer Entstehungsursache ergeben, wenn es nur gelingt, genügend ganaue Instrumente zur Messung der Wellenlänge zu bauen.

Der interessanteste, aber auch subtilste Fall wäre die direkte experimentelle Untersuchung des magnetischen Feldes, welches um einen elektrischen Strom herum entsteht, denn die Erforschung des elastischen Zustandes des Äthers in diesem Falle erlaubten [sic] uns, einen Blick zu werfen in das geheimnisvolle Wesen des elektrischen Stromes. Die Analogie erlaubt uns aber auch sichere Schlüsse über den Ätherzustand im magnetischen Felde, das den elektrischen Strom umgibt, wenn nur die vorher angeführten Untersuchungen zu einem Ziele führen.

Die quantitativen Forschungen über die absoluten Grössen der Dichte und elastischen Kraft des Äthers können, wie ich glaube, erst beginnen, wenn qualitative Resultate existieren, die mit sicheren Vorstellungen verbunden sind; nur eins glaube ich noch sagen zu müssen. Sollte sich die Wellenlänge nicht proportional erweisen $\sqrt{A+k}$, wobei A die elastischen Ätherkräfte a priori, also für uns eine empirisch zu findende Konstante, k die (variable) Stärke des magnetischen Feldes bedeutet, die natürlich den erzeugten in Betracht kommenden elastischen Kräften proportional ist, so wäre der Grund hierfür in der durch die elastische Deformationen erzeugten Veränderung der Dichte des bewegten Äthers zu suchen.

Vor allem aber muss sich zeigen lassen, dass es für den elektrischen Strom zur Bildung des magnetischen Feldes einen passiven Widerstand gibt, der proportional ist der Länge der Strombahn und unabhängig vom Querschnitt und Material des Leiters.