THE COLLECTED PAPERS OF ALBERT EINSTEIN

Volume 2

The Swiss Years:
Writings, 1900-1909

Anna Beck, Translator

Peter Havas, Consultant

Princeton University Press
Princeton, New Jersey

Published by Princeton University Press
41 William Street
Princeton, New Jersey 08540

In the United Kingdom:
Princeton University Press, Chichester, West Sussex

Publication of this translation has been aided by a grant from the National Science Foundation

Printed in the United States of America by
Princeton Academic Press

ISBN: 0-691-08549-8

10 9 8 7 6 5 4 3

CONTENTS

List of Texts

PUBLISHER'S FOREWORD

We are pleased to be publishing this second translation volume of THE COLLECTED PAPERS OF ALBERT EINSTEIN. As with Volume 1, we strongly urge readers to use the translations only together with the documentary edition, which provides the editorial commentary necessary for a more complete understanding of the documents. Every effort has been made to insure the scientific accuracy of this translation. It is not intended as a literary translation that can stand alone without the documentary edition.

We are again grateful to Dr. Anna Beck and Professor Peter Havas for their hard work and dedication to this project, which is separate from the documentary edition project. All translations appearing in the documentary edition were prepared by the editors of that volume, and those appearing in this volume were prepared by Drs. Beck and Havas.

We are also pleased to acknowledge the grant from the National Science Foundation that has made this publication possible. In particular, we thank Dr. Ronald Overman of the NSF for his continued interest in the project.

Princeton University Press
September 1989

PREFACE

This volume contains the translations of all documents in Volume 2 of *The Collected Papers of Albert Einstein*, all of which were originally written in German. It is not self-contained and should be read in conjunction with the documentary edition and its editorial apparatus. All editorial headnotes and footnotes have been omitted, as have the introductory materials and the bibliography. However, we used the bibliography to check the references cited in the documents and, especially, to correct and complete the titles and bibliographic data given by Einstein in his reviews of books and articles. In this volume we have included the editorial footnote numbers, which appear in brackets in the margin and correspond to the footnotes in the documentary edition. We have not corrected any misprints or other errors (including those in the formulas) if the editors have commented on them. Misspellings of names of persons have been routinely corrected.

Although some of the documents have been translated before, we have provided new translations here rather than attempt to use any "best" existing translation.

The purpose of the translation project, in accordance with the agreement between Princeton University Press and the National Science Foundation, is to provide " a careful, accurate translation that is as close to the German original as possible while still producing readable English." This is, therefore, not a "literary" translation but should allow readers who are not fluent in German to make a scholarly evaluation of the content of the documents while also obtaining an appreciation of their flavor.

Many technical expressions used in the original documents are outdated (see the editorial comments in Volume 2); whenever possible, we have not replaced them with the modern English versions but have used the expressions employed in the technical literature of the time, if known, or else we provided a literal translation. In particular, we retained the term "electric mass" frequently used by Einstein for electric charge. All formulas were included in a form as similar to those in the original documents as was possible with our word processor. We kept the standard German notation used at the time, representing vectors by German (Fractur) letters and vector products by [ ]; for example, we kept $[\mathfrak{E}\mathfrak{H}]$ for the vector denoted in current literature by $\vec{E} \times \vec{H}$ or $\mathbf{E} \times \mathbf{H}$.

We are indebted to John Stachel, the Editor of Volume 2, and Robert Schulmann, Associate Editor, as well as Walter Lippincott, Director, and Alice Calaprice, Senior Editor, of Princeton University Press, for their help and encouragement. We also wish to thank Marjorie Zabierek for her part in preparing the final typescript.

Anna Beck, Translator
Peter Havas, Consultant

TEXTS

Doc. 1

CONCLUSIONS DRAWN FROM THE PHENOMENA OF CAPILLARITY

by Albert Einstein

[*Annalen der Physik* 4 (1901): 513-523]

If we denote by γ the amount of mechanical work that we have to supply to a liquid in order to increase the free surface by one unit, then γ is not the total energy increase of the system, as the following cyclic process will show. Let there be a certain amount of liquid of (absolute) temperature T_1 and surface area O_1. We now increase isothermally the surface O_1 to O_2, increase (at constant surface area) the temperature to T_2, then reduce the surface to O_1 and cool the liquid to T_1 again. If one assumes that no heat is supplied to the body other than that received on account of its specific heat, then the total heat supplied to the substance during the cyclic process will be equal to the total heat withdrawn. According to the principle of conservation of energy, the total mechanical work supplied must then also be zero.

Hence the following equation holds:

$$(O_2 - O_1)\gamma_1 - (O_2 - O_1)\gamma_1 = 0 \qquad \text{or} \qquad \gamma_1 = \gamma_2 \, . \qquad [1]$$

However, this contradicts experience. [2]

We have, then, no other choice but to assume that the change in the surface is associated with an exchange of heat as well, and that the surface has a specific heat of its own. If we denote by U the energy, by S the entropy of the unit surface of the liquid, by s the specific heat of the surface, and by w_0 the heat necessary to form a unit surface, expressed in mechanical units, then the quantities

$$dU = s.O.dT + \{\gamma + w_0\} dO$$

and

$$dS = \frac{s.O.dT}{T} + \frac{w_0}{T}\, dO$$

will be total differentials. Hence we will have

$$\frac{\partial(s.O)}{\partial O} = \frac{\partial(\gamma + w_0)}{\partial T} ,$$

$$\frac{\partial}{\partial O}\left[\frac{sO}{T}\right] = \frac{\partial}{\partial T}\left[\frac{w_0}{T}\right] .$$

From these equations it follows that

$$\gamma + w_0 = \gamma - T\frac{\partial\gamma}{\partial T} .$$

This is, however, the total energy necessary to form a unit surface. Further, we form

[3] $$\frac{d}{dT}(\gamma + w_0) = -T\frac{d^2\gamma}{dT^2} .$$

The experimental studies have shown that γ can be represented with
[4] very good approximation as a linear function of temperature, i.e.:

The energy necessary to form a unit surface of a liquid is independent of the temperature.

It also follows that

$$s = \frac{d\gamma}{dT} + \frac{dw_0}{dT} = \frac{d\gamma}{dT} - \frac{d\gamma}{dT} - T\frac{d^2\gamma}{dT^2} = 0 ,$$

[5] hence: no heat content should be ascribed to the surface as such; rather, the energy of the surface is of potential nature. It can be seen already that the quantity

$$\gamma - T\frac{d\gamma}{dT}$$

is more suited for stoichiometric investigations than is the hitherto used γ at boiling temperature. The fact that the energy required for the formation of a unit surface barely varies with the temperature teaches us also that the configuration of molecules in the surface layer will not vary with temperature (apart from changes of the order of magnitude of thermal expansion).

To find a stoichiometric relationship for the quantity

$$\gamma - T\frac{d\gamma}{dT}$$

I proceeded from the simplest assumptions about the nature of molecular attraction forces and examined their consequences regarding their agreement with experiment. In this I was guided by the analogy with gravitational forces. [6]

Let thus the relative potential of two molecules be of the form

$$P = P_\infty - c_1 \cdot c_2 \cdot \varphi(r) \ ,$$

where c is a constant characteristic of the molecule in question, and $\varphi(r)$ is a function of their distance that does not depend on the nature of the molecules. We assume further that [7]

$$\tfrac{1}{2} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} c_\alpha c_\beta \, \varphi(r_{\alpha,\beta})$$

is the corresponding expression for n molecules. In the special case in which all molecules are alike, this expression becomes

$$\tfrac{1}{2} c^2 \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \varphi(r_{\alpha,\beta}) \ .$$

We further make the additional assumption that the potential of the molecular forces has the same magnitude it would have if the matter were homogeneously distributed in space; this is, however, an assumption which we should expect to be only approximately correct. Using it, the above expression converts to

$$P = P_\infty - \tfrac{1}{2} c^2 N^2 \iint d\tau . d\tau' \varphi(r_{d\tau, d\tau'}) \ , \qquad [8]$$

where N is the number of molecules per unit volume. If the molecule of our liquid consists of several atoms, then it shall be possible to put, in analogy with gravitational forces, $c = \Sigma c_\alpha$, where the c_α's denote the values characteristic for the atoms of the elements. If one also puts $1/N = v$, where v denotes the molecular volume, one obtains the final formula

$$P = P_\infty - \tfrac{1}{2}\,\frac{(\Sigma c_\alpha)^2}{v^2} \iint d\tau . d\tau' \varphi(r_{d\tau, d\tau'}) \ .$$

If we now also assume that the density of the liquid is constant up to its surface, which is made plausible by the fact that the energy of the surface is independent of temperature, then we are able to calculate the potential energy per unit volume in the interior of the liquid, and that per unit surface.

I.e., if we put

$$\tfrac{1}{2} \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} dx dy dz . \varphi\left[\sqrt{x^2+y^2+z^2}\right] = K \ ,$$

then the potential energy per unit volume is

$$P = P_\infty - K \frac{(\Sigma c_\alpha)^2}{v^2} \ .$$

If we imagine a liquid of volume V and surface S, we obtain by integration

[9] $$P = P_\infty - K\,\frac{(\Sigma c_\alpha)^2}{v^2} . V - K'\,\frac{(\Sigma c_\alpha)^2}{v^2} . O \ ,$$

where the constant K' denotes

[10] $$\int_{x'=0}^{x'=1} \int_{y'=0}^{y'=1} \int_{z'=-\infty}^{z'=0} \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=\infty} \int_{z=0}^{z=\infty} dx . dy . dz . dx' . dy' . dz' \ \varphi\left[\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}\right] \ .$$

Since nothing is known about φ, we naturally do not get any relationship between K and K'.

One should keep in mind, to begin with, that we cannot know whether or not the molecule of the liquid contains the n-fold mass of the gas molecule, but it follows from our derivation that this does not change our expression for the potential energy of the liquid. Based on the assumptions we have just made, we obtain the following expression for the potential energy of the surface:

$$P = K' \frac{(\Sigma c_\alpha)^2}{v^2} = \gamma - T \frac{d\gamma}{dT} ,$$

or [11]

$$\Sigma c_\alpha = v \sqrt{\gamma - T \frac{d\gamma}{dT}} \cdot \frac{1}{\sqrt{K'}} .$$

Since the quantity on the right can be calculated from R. Schiff's observations for many substances at the boiling temperature, we have ample material for the determination of the quantities c_α. I took all the data from W. Ostwald's book on general chemistry. First, I present here the data [12] that I used for the calculation of c_α for C, H, and O by the least squares method. The column with the heading Σc_α (calc) gives the Σc_α as determined from chemical formulas using the c_α thus obtained. Isomeric compounds were combined into one value, because their values on the left-hand side did not differ significantly from each other. The unit was chosen arbitrarily because it is not possible to determine the absolute value of c_α since K' is unknown.

I found:

$$c_H = -1{,}6, \qquad c_C = 55{,}0, \qquad c_O = 46{,}8 .$$

Formula	Σc_α	Σc_α (calc)	Name of the compound
$C_{10}H_{16}$	510	524	Limonene
CO_2H_2	140	145	Formic acid
$C_2H_4O_2$	193	197	Acetic acid
$C_3H_6O_2$	250	249	Propanoic acid
$C_4H_8O_2$	309	301	Butyric acid and isobutyric acid
$C_5H_{10}O_2$	365	352	Valerianic (pentanoic) acid
$C_4H_6O_3$	350	350	Acetic anhydride
$C_6H_{10}O_4$	505	501	Ethyl oxalate
$C_8H_8O_2$	494	520	Methyl benzoate
$C_9H_{10}O_2$	553	562	Ethyl benzoate
$C_6H_{10}O_3$	471	454	Ethyl-acetoacetate (diacetic ether)
C_7H_8O	422	419	Anisole
$C_8H_{10}O$	479	470	Phenetole and methyl cresolate
$C_8H_{10}O_2$	519	517	Dimethyl resorcinol
$C_5H_4O_2$	345	362	Furfural
$C_5H_{10}O$	348	305	Valeraldehyde
$C_{10}H_{14}O$	587	574	*d*-carvone

It can be seen that in almost all cases the deviations barely exceed the experimental errors and do not show any trend.

After that I separately calculated the values for Cl, Br, and J; these determinations are of course less reliable. I found:

$$c_{\mathrm{Cl}} = 60, \qquad c_{\mathrm{Br}} = 152, \qquad c_{\mathrm{J}} = 198 \ .$$

I present the data in the same way as above:

Formula	Σc_α	Σc_α (calc)	Name of the compound
C_6H_5Cl	385	379	Chloro benzene
C_7H_7Cl	438	434	Chloro toluene
C_7H_7Cl	450	434	Benzyl chloride
C_3H_5OCl	270	270	Epichlorohydrin
C_2OHCl_3	358	335	Chloral
C_7H_5OCl	462	484	Benzoyl chloride
$C_7H_6Cl_2$	492	495	Benzylidene chloride
Br_2	217	304	Bromine
C_2H_5Br	251	254	Ethyl bromide
C_3H_7Br	311	306	Propyl bromide
C_3H_7Br	311	306	Isopropyl bromide
C_3H_5Br	302	309	Allyl bromide
C_4H_5Br	353	354	Isobutyl bromide
$C_5H_{11}Br$	425	410	Isoamyl bromide
C_6H_5Br	411	474	Bromo benzene
C_7H_7Br	421	526	*o*-Bromo toluene
$C_2H_4Br_2$	345	409	Ethylene bromide
$C_3H_6Br_2$	395	461	Propylene bromide
C_2H_5J	288	300	Ethyl iodide
C_3H_7J	343	352	Propyl iodide
C_3H_7J	357	352	Isopropyl iodide
C_3H_5J	338	355	Allyl iodide
C_4H_9J	428	403	Isobutyl iodide
$C_5H_{11}J$	464	455	Isoamyl iodide

It seems to me that the larger deviations from our theory occur for those compounds that have relatively large molecular masses and small molecular volumes.

Based on our assumptions, we found that the expression for the potential energy per unit volume is

$$P = P_\infty - K\frac{(\Sigma c_\alpha)^2}{v^2} \ ,$$

where K denotes a definite quantity, which we, however, are not able to calculate because it is only defined completely by the choice of the c_α's.

We can therefore set $K = 1$ and thereby obtain a definition for the absolute values of the c_α's. If we take this into account from now on, we obtain the following expression for the magnitude of the potential pertaining to one equivalent (molecule):

$$P = P_\infty - K\frac{(\Sigma c_\alpha)^2}{v},$$

where, of course, P_∞ denotes another constant. We could now equate the second member of the right-hand side of this equation to the difference $D_m J - A v_d$, where D_m is the molecular heat of evaporation (heat of evaporation × molecular mass), J the mechanical equivalent of one calorie, A the atmospheric pressure in absolute units, and v_d the molecular volume of the vapor – if the potential energy of the vapor were zero and if at the [13] boiling point the content in kinetic energy would not change during the transition from the liquid to the gaseous state. The first of these assumptions seems to me absolutely safe. However, since we have neither a basis for the second assumption nor a possibility to estimate the quantity in question, we have no other choice but to use the above quantity itself for the calculation.

In the first column of the following table I entered the quantities [14] $\sqrt{D'_m \cdot v}$ in thermal units, with D'_m denoting the heat of evaporation minus the external work of evaporation (in thermal units). In the second column I entered the quantities Σc_α, as obtained from capillarity experiments; the third column contains the quotients of the two values. Isomeric compounds are once again combined into a single line.

Name of the compound	Formula	$\sqrt{D^T_m \cdot v}$	Σc_a (calc)	Quotient
Isobutyl propanoate	$C_7H_{14}0_2$	1157	456	2.54
Isoamyl acetate	"			
Propyl acetate	"			
Isobutyl isobutyrate	$C_8H_{16}0_2$	1257	510	2.47
Propyl valerate	"			
Isobutyl butyrate	"			
Isoamyl propanoate	"			
Isoamyl isobutyrate	$C_9H_{18}0_2$	1367	559	2.45
Isobutyl valerate	"			
Isoamyl valerate	$C_{10}H_{10}0_2$	1464	611	2.51
Benzene	C_6H_6	795	310	2.57
Toluene	C_7H_8	902	372	2.48
Ethyl benzene	C_8H_{10}	1005	424	2.37
m-Xylene	"			
Propyl benzene	C_9H_{12}	1122	475	2.36
Mesitylene	"			
Cymene	$C_{10}H_{14}$	1213	527	2.30
Ethyl formate	$C_3H_60_2$	719	249	2.89
Methyl acetate	"			
Ethyl acetate	$C_4H_80_2$	837	301	2.78
Methyl propanoate	"			
Propyl formate	"			
Methyl isobutyrate	$C_5H_{10}0_2$	882	353	2.50
Isobutyl formate	"			
Ethyl propanoate	"			
Propyl acetate	"			
Methyl butyrate	"			
Ethyl isobutyrate	$C_6H_{12}0_2$	971	405	2.40
Methyl valerate	"			
Isobutyl acetate	"			
Ethyl butyrate	"			
Propyl propanoate	"			
Isoamyl formate	"			

Even though the quotient in the fifth column is by no means a constant, but is, on the contrary, clearly dependent on the constitution of the compounds, we can nevertheless use the material on hand to obtain the factor, or at least its order of magnitude, with which we must multiply our c_a's to obtain them in the absolute unit we had chosen. The mean value of the multiplier looked for is

[15] $$2.51 \times \sqrt{4.17 \times 10^7} = 1.62 \times 10^4 \; .$$

Since the foregoing discussion shows that the kinetic conditions of the molecules change during evaporation (at least if our expression for the potential energy is correct), I decided to obtain the absolute quantity c_α in one more way. I proceeded from the following idea:

If we compress a liquid isothermally and its heat content does not change in the process, as we now wish to assume, then the heat released during [16] compression equals the sum of the work of compression and the work done by the molecular forces. We can therefore calculate the latter work if we can find the amount of heat released during compression. This we can do with the help of Carnot's principle.

Let the state of the liquid be determined by the pressure p in absolute units and by the absolute temperature T; if the value of the heat supplied to the substance during an infinitesimally small change of state is dQ in absolute units, and the mechanical work done on the substance is dA, and if we put

$$dQ = Xdp + S.dT,$$

$$\begin{aligned} dA &= -p.dv = -p\left\{\frac{\partial v}{\partial p}dp + \frac{\partial v}{\partial T}dT\right\} \\ &= p.v.\kappa dp - p.v.\alpha dT \ , \end{aligned} \qquad [17]$$

then the condition that dQ/T and $dQ + dA$ must be total differentials yields the equations

$$\frac{\partial}{\partial T}\left(\frac{X}{T}\right) = \frac{\partial}{\partial p}\left(\frac{S}{T}\right)$$

and

$$\frac{\partial}{\partial T}(X + p\kappa) = \frac{\partial}{\partial p}(S - p\alpha) \ ; \qquad [18]$$

here, as can be seen, X denotes the heat, in mechanical units, supplied to the substance during isothermal compression produced by pressure $p = 1$, S is the specific heat at constant pressure, κ is the coefficient of compressibility, and α is the coefficient of thermal expansion. From these equations, we find

$$Xdp = -T\left\{\alpha + p\frac{\partial \alpha}{\partial p} + p\frac{\partial \kappa}{\partial T}\right\}dp \ . \qquad [19]$$

One has to remember that for any phenomena involving compression of liquids, the atmospheric pressure, to which our bodies are usually subjected, can be considered unhesitatingly as infinitesimally small; likewise, compressions in our experiments are very nearly proportional to the compression forces applied. Thus, the phenomena proceed as if the compression forces were infinitesimally small. If this is taken into account, then our equation reduces to

[20] $$X.dp = -T.\alpha.dp\ .$$

If we now apply the assumption that the kinetic energy of the system does not change in isothermal compression, we obtain the equation

$$X.dp + \text{work of compression} + \text{work of the molecular forces} = 0.$$

If P is the potential of the molecular forces, then the last-mentioned work is

[21] $$\frac{\partial P}{\partial v} \cdot \frac{\partial v}{\partial p} \cdot dp\ .$$

If one inserts herein our expression for the magnitude of the potential of the molecular forces and takes into account that the work of compression is of the order dp^2, one obtains, neglecting this quantity which is infinitesimally small of second order,

[22] $$\frac{T_\alpha}{\kappa} = \frac{(\Sigma c_\alpha)^2}{v^2}$$

where κ denotes the compressibility coefficient in absolute units. We thus obtain once more a means for the determination of the looked-for proportionality coefficient for the quantities c_α. I took the α and κ values for the [23] temperature of ice from Landolt and Börnstein's tables. This yields the following values for the factor sought:

Xylene	1.71×10^4	Ethyl alcohol	1.70×10^4
Cymene	1.71×10^4	Methyl alcohol	1.74×10^4
Turpentine oil	1.73×10^4	Propyl alcohol	1.82×10^4
Ethyl ether	1.70×10^4	Amyl alcohol	2.00×10^4

First of all, it should be noted that the two coefficients obtained by different methods show quite satisfactory agreement even though they have been derived from totally different phenomena. The last table shows a very satisfactory agreement of the values; only the higher alcohols show deviations. [24] This is to be expected, because from the deviations of alcohols from [25] Mendeleev's thermal expansion law and from R. Schiff's stoichiometric law of [26] capillarity, it has already been concluded earlier that in these compounds temperature changes are associated with changes in the size of the molecules of the liquid. Hence it is to be expected that such molecular changes should also arise during isothermal compression, so that for such compounds at the same temperature the heat content will be a function of volume.

In summary, we may state that our basic assumption stood the test: To each atom corresponds a molecular attraction field that is independent of the temperature and of the way in which the atom is chemically bound to other atoms.

Finally, it should also be pointed out that the constants c_α generally increase with increased atomic weight, but not always, and not in a proportional way. The question of whether and how our forces are related to gravitational forces must therefore be left completely open for the time being. It [27] should also be added that the introduction of the function $\varphi(r)$, which is taken to be independent of the nature of the molecules, should be understood as an approximate assumption, and so should the replacement of sums by integrals; [28] in fact, as the example of water shows, our theory does not seem to [29] hold for substances with small atomic volumes. Only extensive special investigations can be expected to bring answers to these questions.

Zurich, 13 December 1900. (Received on 16 December 1900)

Doc. 2

ON THE THERMODYNAMIC THEORY OF THE DIFFERENCE IN POTENTIALS BETWEEN METALS AND FULLY DISSOCIATED SOLUTIONS OF THEIR SALTS AND ON AN ELECTRICAL METHOD FOR INVESTIGATING MOLECULAR FORCES

By A. Einstein

[*Annalen der Physik* 8 (1902): 798-814]

§1. A hypothetical extension of the second law of the mechanical theory of heat

The second law of the mechanical theory of heat can be applied to such physical systems which are capable of passing, with any desired approximation, through reversible cyclic processes. In accordance with the derivation of this law from the impossibility of converting latent heat into mechanical energy, it is here necessary to assume that those processes are realizable. However, in an important application of the mechanical theory of heat, namely the mixing of two or more gases by means of semipermeable membranes, it is doubtful whether this postulate is satisfied. The thermodynamic theory of dissociation of gases and the theory of dilute solutions are based on the assumption that this process is realizable. [1]

As is well known, the assumption to be introduced is as follows: For any two gases A and B it should be possible to produce two partitions such that one is permeable for A but not for B, while the other is permeable for B but not for A. If the mixture consists of more than two components, then this assumption becomes even more complicated and improbable. Since the results of the theory have been completely confirmed by experiment despite the fact that we worked with processes whose realizability could indeed be doubted, the question arises whether the second law could not be applied to ideal processes of a certain kind without contradicting experience.

In this sense, on the basis of the experience obtained, we certainly can advance the proposition: One remains in agreement with experience if one extends the second law to physical mixtures whose individual components are restricted to certain subspaces by conservative forces acting in certain planes. We shall hypothetically generalize this proposition to the following:

One remains in agreement with experience when one applies the second law to physical mixtures whose individual components are acted upon by arbitrary conservative forces. [2]

In the following we will always make use of this hypothesis, even when this does not seem absolutely necessary.

§2. *On the dependence of the electric potential difference of a completely dissociated salt solution and an electrode consisting of the solute metal on the concentration of the solution and the hydrostatic pressure*

Let a solution of a completely dissociated salt be contained in a cylindrical vessel whose axis coincides with the z-axis of a Cartesian coordinate system. Let νdo be the number of gram-molecules of the salt dissolved in the volume element do, $\nu_m do$ the number of metal ions, and [3] $\nu_s do$ the number of acid ions, where ν_m and ν_s are integral multiples of ν, so that we have the following equations:

$$\nu_m = n_m \cdot \nu \ ,$$
$$\nu_s = n_s \cdot \nu \ .$$

Further, let $n.\nu.E.do$ be the magnitude of the total positive electric charge of the ions in do, and hence also, up to the infinitesimally small, the magnitude of the negative charge. Here n is the sum of valencies of the molecule's metal ions, and E the amount of electricity required for the electrolytic separation of one gram-molecule of a univalent ion. [4]

These equations are certainly valid, since the number of excess ions of one kind can be neglected.

We shall further assume that the metal and acid ions are acted upon by an external conservative force whose potential per ion has the magnitude P_m [5] and P_s, respectively. Furthermore, we neglect the variability of the density of the solvent with the pressure and density of the dissolved salt, and assume that a conservative force, whose potential per gram-equivalent of the solvent has the magnitude P_0, acts upon the parts of the solvent; there shall be $\nu_0 do$ gram-molecules of solvent in do.

Suppose that all force functions depend solely on the z-coordinate, and that the system is in electrical, thermal, and mechanical equilibrium. Then the quantities concentration ν, electric potential π, osmotic pressures of the two ion types p_m and p_s, and hydrostatic pressure p_0 will be functions of z only.

At each location of the electrolyte, each of the two types of electrons must then be in equilibrium separately, which is expressed by the equations

$$-\frac{dp_m}{dz}\cdot\frac{1}{\nu}-n_m\frac{dP_m}{dz}-nE\frac{d\pi}{dz}=0\ ,$$
$$-\frac{dp_s}{dz}\cdot\frac{1}{\nu}-n_s\frac{dP_s}{dz}+nE\frac{d\pi}{dz}=0\ ,$$

where

$$p_m=\nu\cdot n_m\cdot RT\ ,$$
$$p_s=\nu\cdot n_s\cdot RT\ ,$$

and where R is a constant common to all ionic species. Hence the equations take the form

$$(1)\qquad \left\{\begin{aligned} &n_mRT\frac{d\lg\nu}{dz}+n_m\frac{dP_m}{dz}+nE\frac{d\pi}{dz}=0\ ,\\ &n_sRT\frac{d\lg\nu}{dz}+n_s\frac{dP_s}{dz}-nE\frac{d\pi}{dz}=0\ .\end{aligned}\right.$$

If P_m and P_s are known for all z, and ν and π for a particular z, then equations (1) yield ν and π as functions of z. Also, the condition that the solution as a whole is in equilibrium would result in an equation for the determination of the hydrostatic pressure p_0, which need not be written [6] down. We only note that the reason that dp_0 is independent of $d\nu$ and $d\pi$ is that we are free to postulate arbitrary conservative forces that act on the molecules of the solvent.

We now imagine that electrodes made of the solute metal and occupying a vanishingly small part of the cross section of the cylindric vessel are placed in the solution at $z = z_1$ and $z = z_2$. The solution and the electrodes together form a physical system, which we take through the following reversible isothermal cyclic process:

1st partial process: We pass the amount of electricity nE *infinitely slowly* through the solution, using the electrode at $z = z_1$ as anode, and that at $z = z_2$ as cathode.

2nd partial process: The amount of dissolved metal that has thus been moved electrolytically from z_1 to z_2 we now move back mechanically infinitely slowly from z_2 to z_1.

First of all, it is evident that the process is strictly reversible, since all steps are imagined to proceed infinitely slowly, i.e., the process is compounded of (ideal) states of equilibrium. For such a process the second law requires that the total amount of heat supplied to the system during the cyclic process shall vanish. In conjunction with the second law, the first law requires that the sum of all other energies supplied to the system during the cyclic process shall vanish.

During the first partial process the amount of electric work supplied is

$$-nE(\Pi_2 - \Pi_1) \ ,$$

where Π_2 *and* Π_1 denote the electric potentials of the electrodes.

During the second partial process

$$\int_{z_2}^{z_1} K dz$$

is supplied, where K is the force acting in the positive z-direction that is required for the n_m metal ions that are to be moved, and which are now in the metallic state, to keep them at rest at an arbitrary location z. It is easily seen that the following equation will hold for K:

$$K - n_m \frac{dP_m}{dz} - n_m v_m \frac{dp_o}{dz} = 0 \ .$$

Here v_m denotes the volume of one metal ion in the metallic state. Hence the above work takes on the value

$$\int_{z_2}^{z_1} K.dz = - \int_{z_1}^{z_2} \left[n_m \frac{dP_m}{dz} + n_m v_m \frac{dp_0}{dz} \right] dz$$

$$= -n_m [(P_{m_2} - P_{m_1}) + v_m (p_{0_2} - p_{0_1})] \ ,$$

where the second index denotes the coordinate of the electrode.

We obtain, hence, the equation

$$(2) \qquad n.E.(\Pi_2 - \Pi_1) = -n_m(P_{m_2} - P_{m_1}) - n_m v_m(p_{0_2} - p_{0_1}) \ .$$

If the electric potentials in the cross sections of the electrodes inside the solution are denoted by π_1 and π_2, integration of the first equation (1) yields

$$- n.E(\pi_2 - \pi_1) = n_m[P_{m_2} - P_{m_1}] + n_m RT \log\left[\frac{\nu_2}{\nu_1}\right] \ ,$$

where ν_1 and ν_2 refer again to the cross sections of the electrodes. Adding these equations, one obtains

$$(3) \qquad \begin{cases} (\Pi_2 - \pi_2) - (\Pi_1 - \pi_1) = (\Delta\Pi)_2 - (\Delta\Pi)_1 \\ \qquad = \dfrac{n_m RT}{nE}\log\left[\dfrac{\nu_2}{\nu_1}\right] - \dfrac{n_m v_m}{nE}(p_{0_2} - p_{0_1}) \ . \end{cases}$$

Since the ν's and p_0 are completely independent of each other, this equation represents the dependence of the potential difference $\Delta\Pi$ between metal and solution on concentration and hydrostatic pressure. It should be noted that the postulated forces no longer appear in the result. If they were to appear, the hypothesis posited in §1 would have been carried ad absurdum. The equation obtained can be resolved into two equations, namely:

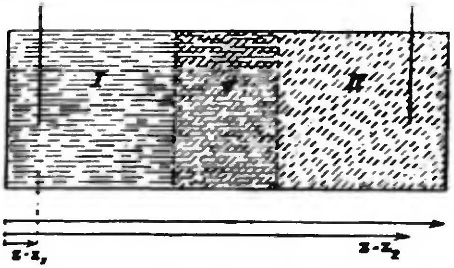
$$(4) \qquad \begin{cases} (\Delta\Pi)_2 - (\Delta\Pi)_1 = \dfrac{n_m}{n} \dots \dfrac{RT}{E}\log\left[\dfrac{\nu_2}{\nu_1}\right] & \text{at constant pressure,} \\ (\Delta\Pi)_\pi - (\Delta\Pi)_1 = - \dfrac{n_m}{n}\dfrac{v_m}{E} \dots (p_{0_2} - p_{0_1}) & \text{at constant concentration.} \end{cases}$$

The final formula (3) could have also been obtained without the hypothesis proposed in §1 had the external forces been identified with terrestrial gravity. However, in that case v and p would not be independent of each other and the resolution into equations (4) would not be permitted.

It should also be briefly noted that the Nernst theory of electric forces inside dissociated electrolytes, taken in conjunction with the first of equations (4), makes it possible to calculate the electromotive force of the concentration cell. Thus one arrives at a result that has already been tested repeatedly and that till now has been derived from special assumptions. [7]

§3. *On the dependence of the quantity* $\Delta\Pi$ *on the nature of the acid*

We shall consider the following ideal state of equilibrium: Let us again have a cylindric vessel. Parts *I* and *II* shall each contain a completely dissociated salt solution; the two salts shall have an identical metal ion (same metal and same electric charge) but a different acid ion. Between the two parts there shall be a connecting space V which contains both salts in solution. Upon the acid ions in V shall act forces whose potentials $P_s^{(1)}$ and $P_s^{(2)}$ depend only on z, and these forces shall bring about that only infinitesimally few acid ions of the first and of the second type get into *II* and *I*, respectively. Furthermore, $P_s^{(1)}$ and $P_s^{(2)}$ shall be chosen such that the metal ion concentration in the two parts *I* and *II* be the same. Also, let $p_{0_1} = p_{0_2}$.



If there are per unit volume $\nu_m^{(1)}$ and $\nu_m^{(2)}$ metal ions that correspond to the first and second type of salt, respectively, then

$$\nu_{m_1}^{(1)} = \nu_{m_2}^{(2)}, \quad \nu_{s_1}^{(2)} = 0, \quad \nu_{s_2}^{(1)} = 0 , \tag{1}$$

where the subscripts refer to space *I* and *II*, respectively.

However, the condition for the equilibrium of the metal ions in V yields

$$-RT\,\frac{d\log(\nu_m^{(1)} + \nu_m^{(2)})}{dz} - \epsilon E\,\frac{d\pi}{dz} = 0\ ,$$

where ϵ denotes the valency of the metal ion.

Integrating over V and taking equations (1) into account, we obtain

$$\pi_2 = \pi_1\ . \tag{2}$$

Next we imagine that electrodes made of the solute metal are installed in *I* and *II*, and construct the following ideal cyclic process:

1st partial process: We send an amount of electricity ϵE infinitely slowly through the system, taking the electrode in *I* as anode, and the other as cathode.

2nd partial process: The metal thus transported electrolytically from $z = z_1$ to $z = z_2$, which has the mass of one gram-equivalent, is now returned mechanically to the electrode in $z = z_1$.

By applying the two laws of the mechanical theory of heat, one again reaches the conclusion that the sum of mechanical and electrical energy supplied to the system during the cyclic process vanishes. Since, as one can readily see, the second step does not require any energy, one obtains the equation

$$\Pi_2 = \Pi_1\ , \tag{3}$$

where Π_2 and Π_1 again denote the potentials of the electrodes. By subtracting equations (3) and (2), one obtains

$$(\Pi_2 - \pi_2) - (\Pi_1 - \pi_1) = (\Delta\Pi)_2 - (\Delta\Pi)_1 = 0$$

and hence the following theorem:

The potential difference between a metal and a completely dissociated solution of a salt of this metal in a given solvent is independent of the nature of the electronegative component, and depends solely on the concentration of the metal ions. It is assumed, however, that the metal ion of these salts is charged with the same amount of electricity.

§4.

Before we turn to the study of the dependency of $(\Delta\Pi)$ on the nature of the solvent, we shall briefly develop the theory of conservative molecular forces in liquids. I shall borrow the notation from a previous article on this topic,[1] which shall at the same time temporarily justify the hypotheses I am going to introduce.

To each molecule of a liquid or a substance dissolved in a liquid shall be assigned a certain constant c, so that the expression for the relative potential of molecular forces of two molecules, which shall be characterized by the indices $\ldots_1$ and $\ldots_2$, will be

(a) $$P = P_\infty - c_1 c_2 \varphi(r) ,$$

where $\varphi(r)$ is a function of distance common to all molecular species. These forces shall simply superpose, so that the expression for the relative potential of n molecules shall have the form

(b) $$\text{Const.} - \tfrac{1}{2} \sum_{\alpha=1}^{\alpha=n} \sum_{\beta=1}^{\beta=n} c_\alpha c_\beta \, \varphi(r_{\alpha\beta})$$

Should all molecules be identical, we would obtain the expression

(c) $$\text{Const.} - \tfrac{1}{2} c^2 \sum_{\alpha=1}^{\alpha=n} \sum_{\beta=1}^{\beta=n} \varphi(r_{\alpha\beta}) .$$

Further, if the laws of interaction and distribution of the molecules are so constituted that it is permissible to convert the sums into integrals, then this expression becomes

$$\text{Const.} - \tfrac{1}{2} c^2 N^2 \int\int d\tau . d\tau' \varphi(r_{d\tau, d\tau'}) .$$

---

[1] A. Einstein, *Ann. d. Physik* 4 (1901): 513. [9]

Here N denotes the number of molecules per unit volume. If N_0 denotes the number of molecules in one gram-equivalent, then $N_0/N = v$ is the molecular [8] volume of the liquid, and if we assume that the investigation involves one gram-equivalent and neglect the effect of the liquid surface, our expression becomes

$$\text{Const.} - \tfrac{1}{2}\,\frac{c^2}{v}\,N_0^2 \int_{-\infty}^{\infty} d\tau'.\varphi(r_{0,d\tau'}) \;.$$

We shall now choose the unit for c such that this expression reduces to

$$\text{(d)} \qquad \text{Const.} - \frac{c^2}{v}, \text{ hence } \tfrac{1}{2}\,N_0^2 \int_{-\infty}^{\infty} d\tau'.\varphi(r_{0,d\tau'}) = 1 \;.$$

By this choice one obtains absolute units for the quantities c. It has been shown in the previously cited article that one remains in agreement with experience if one sets $c = \Sigma c_\alpha$, where the quantities c_α refer to the atoms composing the molecule.

We now want to calculate the relative attraction potential of a gram-molecule of an ion with respect to its solvent, while making the express assumption that the attraction fields of the solvent molecules do not act upon the electric charges of the ions. Methods to be developed later will provide the means by which to decide whether this assumption is permissible.

If c_j is the molecular constant of the ion and c_ℓ that of the solvent, then the potential of one molecule of the ion with respect to the solvent has the form

$$\text{Const.} - \sum_{\ell} c_j.c_\ell.\varphi(r) = \text{const.} - c_j.c_\ell N_\ell \int d\tau.\varphi(r_{0,d\tau}) \;,$$

where N_ℓ denotes the number of solvent molecules per unit volume. Since $N_0/N_\ell = v_\ell$, this expression becomes

$$\text{Const.} - c_j.c_\ell.\frac{N_0}{v_\ell} \int d\tau.\varphi(r_{0,d\tau}) \;.$$

However, since a gram-equivalent contains N_0 molecules of the ion, we obtain for the relative potential of one gram-equivalent of the ion:

$$\text{Const.} - \frac{c_j \cdot c_\ell}{v_\ell} N_0^2 \int d\tau . \varphi(r_{0,d\tau}) = \text{const.} - 2\frac{c_j \cdot c_\ell}{v_\ell} .$$

Introducing the solvent concentration $1/v_\ell = \nu_\ell$, one obtains the form

$$\text{(e)} \qquad P_{j\ell} = \text{const.} - 2c_j \cdot c_\ell \nu_\ell .$$

If the solvent is a mixture of several liquids, which we shall distinguish from each other by indices, we obtain

$$\text{(e}'\text{)} \qquad P_{j\ell} = \text{const.} - 2c_j \sum c_\ell \nu_\ell ,$$

where the ν_ℓ denote the number of gram-molecules of the individual components of the solvent per unit volume. The formula (e') holds approximately also if the quantities ν_ℓ vary with position.

§5. *On the dependence of the electric potential difference existing between a metal and a completely dissociated solution of a salt of this metal on the nature of the solvent*

Let a cylindric vessel again be divided, as in §3, into spaces *I* and *II* and the connecting space *V*. Space *I* shall contain a first solvent, *II* a second one, and *V* a mixture of both, and forces that prevent diffusion shall act on the solvents in space *V*. The vessel shall contain a completely dissociated dissolved salt. In *V*, on its anions there shall act forces whose potential shall be called P_s and which shall be chosen such that the salt be of the same concentration in *I* and *II*. We now establish the condition for the equilibrium of the metal ions. We again take the z-axis parallel to the cylinder axis from *I* to *II*.

The force of electric origin that acts on one gram-equivalent will be

$$- \frac{n}{n_m} E \frac{d\pi}{dz} .$$

The force exerted on the equivalent by osmotic pressure is

$$- RT \frac{d \log \nu}{dz} .$$

The effect of molecular forces on the equivalent is

$$- \frac{d}{dz}\{- 2c_m c_\ell^{(1)} \nu_\ell^{(1)} - 2c_m c_\ell^{(2)} \nu_\ell^{(2)}\} ,$$

where the superscripts refer to the solvents. The equilibrium condition sought is then

$$- \frac{n}{n_m} E \frac{d\pi}{dz} - RT \frac{d \log \nu}{dz} + \frac{d}{dz}\{2c_m c_\ell^{(1)} \nu_\ell^{(1)} + 2c_m c_\ell^{(2)} \nu_\ell^{(2)}\} = 0 .$$

If one integrates over V and takes into account that ν is the same in *I* and *II*, and that according to our assumption $\nu_\ell^{(1)}$ and $\nu_\ell^{(2)}$ vanish, one obtains

$$\pi_2 - \pi_1 = \frac{n_m}{n} \frac{2c_m}{E} \{c_\ell^{(2)} \nu_\ell^{(2)} - c_\ell^{(1)} \nu_\ell^{(1)}\} ,$$

where the superscripts refer to spaces *I* and *II*, respectively.

We now imagine that electrodes made up of the dissolved metal are placed into *I* and *II*, and construct a cyclic process by sending an amount of electricity $\frac{n}{n_m} E$ through the system and then returning the transported metal mechanically, which does not require any work if we assume that the hydrostatic pressure is the same in *I* and *II*. Application of the two laws of the theory of heat yields

$$\Pi_2 - \Pi_1 = 0 .$$

Subtraction of the two results gives

$$\begin{aligned}(\Pi_2 - \pi_2) - (\Pi_1 - \pi_1) &= (\Delta\Pi)^{(2)} - (\Delta\Pi)^{(1)} \\ &= - \frac{n_m}{n} \frac{2c_m}{E} \{c_\ell^{(2)} \nu_\ell^{(2)} - c_\ell^{(1)} \nu_\ell^{(1)}\} .\end{aligned}$$

If each of the two solvents is a mixture of several nonconducting liquids, one obtains somewhat more generally

$$(\Delta\Pi)^{(2)} - (\Delta\Pi)^{(1)} = -\frac{n_m}{n}\frac{2c_m}{E}\left\{\sum c_\ell^{(2)}\nu_\ell^{(2)} - \sum c_\ell^{(1)}\nu_\ell^{(1)}\right\} ,$$

where now ν_ℓ denotes the number of gram-molecules of a component of the solvent in a volume element of the mixed solvent.

Hence the potential difference $\Delta\Pi$ depends on the nature of the solvent. This dependence can be used as a basis for a method of exploring the molecular forces.

§6. *A method for the determination of the constant* c *for metal ions and solvents*

Let two completely dissociated salt solutions undergo diffusion in a cylindrical vessel; these salts shall be indicated by subscripts. The solvent shall be the same throughout the vessel and shall be indicated by the superscript. The vessel shall again be divided into spaces *I* and *II* and the connecting space *V*. Space *I* shall contain only the first salt, and *II* only the second salt; diffusion of the two salts shall take place in space *V*. Into spaces *I* and *II* there shall be introduced electrodes consisting of the respective metal solute and having electric potentials Π_1 and Π_2', respectively; onto the second electrode shall be soldered a piece of the first electrode metal, whose potential is Π_2. Furthermore, we denote the electric potentials in the interior of the unmixed solutions in *I* and *II* by π_1 and π_2. We then have

$$(\Pi_2 - \Pi_1)^{(1)} = (\Pi_2 - \Pi_2') + (\Pi_2' - \pi_2)^{(1)} + (\pi_2 - \pi_1)^{(1)} - (\Pi_1 - \pi_1)^{(1)} .$$

If one produces exactly the same arrangement except for using a different solvent, which shall be denoted by the superscript $^{(2)}$, one obtains:

$$(\Pi_2 - \Pi_1)^{(2)} = (\Pi_2 - \Pi_2') + (\Pi_2' - \pi_2)^{(2)} + (\pi_2 - \pi_1)^{(2)} - (\Pi_1 - \pi_1)^{(2)}$$

Subtracting these two expressions and taking into account the results found in §5, one obtains

$$(\Pi_2 - \Pi_1)^{(2)} - (\Pi_2 - \Pi_1)^{(1)} =$$
$$\{(\pi_2 - \pi_1)^{(2)} - (\pi_2 - \pi_1)^{(1)}\} - \frac{2}{E}\left\{\left[\frac{c_m n_m}{n}\right]_2 - \left[\frac{c_m n_m}{n}\right]_1\right\} \cdot \{c_\ell^{(2)} \nu_\ell^{(2)} - c_\ell^{(1)} \nu_\ell^{(1)}\} \ .$$

The extension required if the solvents are mixtures is easily obtained as in §5.

The values of the left-hand side of this equation are obtained directly from experiment. The determination of the first term of the right-hand side will be dealt with in the next paragraph; for the time being, let it only be said that this term can be calculated from the concentrations used and the molecular conductivities of the respective ions for the respective solvent, provided the arrangement has been suitably chosen. Thus the equation makes it possible to calculate the second term on the right-hand side.

This we utilize to determine the constant c for the metal ions and to test our hypotheses. We always use the same two solvents in a series of experiments of the kind described. Then for the whole experimental series the quantity

$$\frac{2}{E}\{c_\ell^{(2)} v_\ell^{(2)} - c_\ell^{(1)} v_\ell^{(1)}\} = k = \text{const.}$$

[10] Hence, if one puts $n_1/n_{m_1} = E_1$, etc., to be equal the valency of the first etc. metal ion, the last term calculated of the right-hand side will be a relative measure for the quantity

$$\left[\frac{c_{m_2}}{\epsilon_2} - \frac{c_{m_1}}{\epsilon_1}\right] .$$

If one thus examines the combinations of all electrode metals pair by pair, one obtains the quantities

$$\left\{\frac{c_{m_j}}{\epsilon_j} - \frac{c_{m_k}}{\epsilon_k}\right\}$$

in relative measure.

One obtains in this same measure the quantities c_m/ϵ separately by carrying out an analogous investigation with a metal in such a way that the salts and electrodes in *I* and *II* contain the same metal, but that ϵ, i.e., the valency (electrical charge) of the metal ion, is different on the two sides. The value of the quantities c_m in this measure can then be obtained for the individual metals. A series of such experiments thus leads to the ratios of the c_m's, i.e., the constants for the molecular attraction of metal ions. This series of c_m's must be independent of the nature of the salts used, and the ratios of the c_m's thus obtained must be independent of the nature of the two solvents on which we based the investigation. A further requirement must be that c_m must prove to be independent of the electrical charge (valency) displayed by the ion. If this is the case, the above assumption that the molecular forces do not act upon the electrical charges is correct.

If one wishes to determine the absolute value of the quantities c_m at least approximately, one can do so by taking the approximate value of k for both solvents from the results of the previously cited paper using the formula $c = \Sigma c_\alpha$. It has to be noted here, of course, that just for the two liquids most obviously suggesting themselves as solvents, namely water and alcohol, it has not been possible to demonstrate the validity of the law of attraction from the phenomena of capillarity, evaporation, and compressibility. [11]

Our results could equally well serve as a basis for studying the solvent constants c_ℓ, however, by basing the investigation on two metal ions and varying the solvent, so that then the quantity

$$\frac{2}{E}\left\{\left[\frac{c_m n_m}{n}\right]_2 - \left[\frac{c_m n_m}{n}\right]_1\right\}$$

is to be considered as constant. By also using mixtures for solvents, the investigation might be extended to all electrically nonconductive liquids. From such experiments it is possible to calculate relative values of the quantities c_a that pertain to the atoms constituting the liquid molecules. This, too, opens ample possibilities for testing the theory inasmuch as the c_a can be arbitrarily overdetermined. Here, too, the result must be independent of the choice of the metal ions.

§7. *Calculation of* $(\pi_2 - \pi_1)$

All that now remains is to study the diffusion process in the space V in greater detail. Let the variable quantities depend on z only, where the z-axis of the Cartesian coordinate system we have chosen coincides with the direction of the axis of our vessel. ν_{m_1}, ν_{s_1}, ν_{m_2}, and ν_{s_2} shall be the z-dependent concentrations (gram-equivalents per unit volume) of the four ionic species, $\epsilon_{m_1}E$, $-\epsilon_{s_1}E$, $\epsilon_{m_2}E$, $-\epsilon_{s_2}E$ their electric charges, and π the electric potential. Since no substantial electric charges occur anywhere, we have for all z approximately

$$(\alpha) \qquad \nu_{m_1}\epsilon_{m_1} - \nu_{s_1}\epsilon_{s_1} + \nu_{m_2}\epsilon_{m_2} - \nu_{s_2}\epsilon_{s_2} = 0 .$$

In addition, for each ionic species we obtain an equation which states that the increase per unit time in the number of ions of a certain kind present in one volume element equals the difference between the number of molecules entering and the number of molecules leaving that volume element during the same time period:

[12]
$$(\beta) \qquad \left\{\begin{array}{l} v_{m_1}\cdot\frac{\partial}{\partial z}\left\{RT\frac{\partial \nu_{m_1}}{\partial z} + \epsilon_{m_1}\nu_{m_1}E\,\frac{\partial \pi}{\partial z}\right\} = \frac{\partial \nu_{m_1}}{\partial t} , \\ \\ v_{s_1}\cdot\frac{\partial}{\partial z}\left\{RT\frac{\partial \nu_{s_1}}{\partial z} + \epsilon_{s_1}\nu_{s_1}E\,\frac{\partial \pi}{\partial z}\right\} = \frac{\partial \nu_{s_1}}{\partial t} , \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right.$$

where v with the corresponding subscript denotes the constant velocity imparted by a unit mechanical force to one gram-equivalent of the corresponding ion in the solution.

In conjunction with the boundary conditions, these four equations completely determine the process taking place, since they permit the determination of the five quantities

$$\frac{\partial \pi}{\partial z}, \frac{\partial \nu_{m_1}}{\partial t} \dots \frac{\partial \nu_{s_2}}{\partial t}$$

uniquely for all times. The general treatment of the problem would entail great difficulties, however, especially since equations (β) are not linear in the unknowns. However, we are only interested in the determination of $\pi_2 - \pi_1$. We therefore multiply the equations (β) successively by ϵ_{m_1}, $-\epsilon_{s_1}$, ϵ_{m_2}, $-\epsilon_{s_2}$, and obtain, when taking into account (α),

$$\frac{\partial \varphi}{\partial z} = 0 \ ,$$

where

$$\varphi = RT\left\{ v_{m_1}\epsilon_{m_1}\frac{\partial \nu_{m_1}}{\partial z} - v_{s_1}\epsilon_{s_1}\frac{\partial \nu_{s_1}}{\partial z} + \cdot \cdot \cdot \right\} + \{ v_{m_1}\epsilon^2_{m_1}\nu_{m_1} + v_{s_1}\epsilon^2_{s_1}\nu_{s_1} + \cdot + \cdot \} \frac{\partial \pi}{\partial z} \ . \qquad [13]$$

In view of the fact that

$$\frac{\partial \nu_{m_1}}{\partial z}, \frac{\partial \nu_{s_1}}{\partial z} \dots \frac{\partial \pi}{\partial z}$$

vanish wherever diffusion does not take place, integration of this equation with respect to z yields

$$\varphi = 0 \ .$$

Since time is to be considered as constant, we may write

$$d\pi = -\frac{RT\{ v_{m_1}\epsilon_{m_1}d\nu_{m_1} - v_{s_1}.\epsilon_{s_1}.d\nu_{s_1} + v_{m_2}\epsilon_{m_2}d\nu_{m_2} - v_{s_2}\epsilon_{s_2}d\nu_{s_2}\}}{v_{m_1}\epsilon^2_{m_1}\nu_{m_1} + v_{s_1}\epsilon^2_{s_1}\nu_{s_1} + v_{m_2}\epsilon^2_{m_2}\nu_{m_2} + v_{s_2}.\epsilon^2_{s_2}\nu_{s_2}} \ . \qquad [14]$$

In general, the expression on the right is not a total differential, which means that $\Delta\Pi$ is determined not only by the concentrations in diffusion-free regions but also by the character of the diffusion process. However, one can make the integration possible by applying an artifice in the arrangement.

We imagine that space V is divided into three parts, space (1), space (2), and space (3), and that these are separated from each other by two partitions before the start of the experiment. Let (1) be connected with *I* and (3) with *II*, and let the two salts be simultaneously dissolved in (2), at concentrations that shall be exactly the same as in *I* and *II*, respectively. Thus, before the experiment, (1) and *I* contain only the first salt in solution, *II* and (3) only the second, and (2) a mixture of both. The concentration is everywhere constant. At the start of the experiment the partitions are removed and immediately thereafter the potential difference between the two electrodes is measured. For this time it is possible to integrate over the diffusing layers, because ν_{m_1} and ν_{s_1} in the first diffusing layer, and ν_{m_2} and ν_{s_2} in the second, are constant. The integration yields

$$\pi_2 - \pi_1 = RT\Bigg\{\frac{v_{m_1} - v_{s_1}}{v_{m_1}\epsilon_{m_1} + v_{s_1}\epsilon_{s_1}} \lg\left[1 + \frac{v_{m_1}\epsilon^2_{m_1}\nu_{m_1} + v_{s_1}\epsilon^2_{s_1}\nu_{s_1}}{v_{m_2}\epsilon^2_{m_2}\nu_{m_2} + v_{s_2}\epsilon^2_{s_2}\nu_{s_2}}\right]$$

$$- \frac{v_{m_2} - v_{m_2}}{v_{m_2}\epsilon_{m_2} + v_{s_2}\epsilon_{s_2}} \lg\left[1 + \frac{v_{m_2}\epsilon^2_{m_2}\nu_{m_1} + v_{s_2}\epsilon^2_{s_2}\nu_{s_2}}{v_{m_1}\epsilon^2_{m_1}\nu_{m_1} + v_{s_1}\epsilon^2_{s_1}\nu_{s_1}}\right]\Bigg\} .$$

The method can be simplified if it is possible to choose the same acid ion of the same concentration in *I* and *II*. If in this case *I* is connected directly with space *II*, one has to put for the start of the diffusion process:

$$\frac{\partial(\nu_{s_1} + \nu_{s_2})}{\partial z} = 0; \quad \nu_{s_1} + \nu_{s_2} = \nu_s = \text{const.}$$

Similarly, according to assumption:

$$\epsilon_{s_1} = \epsilon_{s_2} = \epsilon_s \quad \text{and} \quad v_{s_1} = v_{s_2} = v_s .$$

Equation (1) then becomes [15]

$$\nu_{m_1}\epsilon_{m_1} + \nu_{m_2}\epsilon_{m_2} - \nu_s\epsilon_s = 0 \,. \tag{1'}$$

Of the equations (2), the first and the third remain unchanged, and the second and the fourth yield, by addition,

$$v_s \frac{\partial}{\partial z}\left\{RT\frac{\partial \nu_s}{\partial z} - \epsilon_s \nu_s E \frac{\partial \pi}{\partial z}\right\} = \frac{\partial \nu_s}{\partial t} \,.$$

If the derivatives with respect to time are eliminated by means of equation (1') from the equations (2) thus modified, one obtains, as previously, an expression for $d\pi$, that is a total differential. Integrating, one gets

$$\pi_2 - \pi_1 = -\frac{RT}{E}\,\frac{v_{m_2} - v_{m_1}}{v_{m_2}\epsilon_2 - v_{m_1}\epsilon_1}\,\lg\frac{\epsilon_{m_2}^2 v_{m_2}\nu_{m_2} + \epsilon_s^2 v_s\nu_s}{\epsilon_{m_1}^2 v_{m_1}\nu_{m_1} + \epsilon_s^2 v_s\nu_s} \,,$$

where the numerical indices now refer to the integration limits. Due to the relations

$$\epsilon_{m_1}\nu_{m_1} = \epsilon_s\nu_s = \epsilon_{m_2}\nu_{m_2} \,,$$

we obtain even more simply

$$\pi_2 - \pi_1 = -\frac{RT}{E}\,\frac{v_{m_2} - v_{m_1}}{v_{m_2}\epsilon_2 - v_{m_1}\epsilon_1}\,\lg\frac{\epsilon_{m_2} v_{m_2} + \epsilon_s v_s}{\epsilon_{m_1} v_{m_1} + \epsilon_s v_s} \,.$$

In conclusion, I feel the need to apologize for outlining here a skimpy plan for a laborious investigation without contributing anything to its experimental solution; but I am not in the position to do so. All the same, this work will have achieved its goal if it motivates a researcher to tackle the problem of molecular forces from this direction. [16]

Bern, April 1902. (Received on 30 April 1902)

Doc. 3

KINETIC THEORY OF THERMAL EQUILIBRIUM AND OF THE SECOND LAW OF THERMODYNAMICS

by A. Einstein

[*Annalen der Physik* 9 (1902): 417-433]

Great as the achievements of the kinetic theory of heat have been in the domain of gas theory, the science of mechanics has not yet been able to produce an adequate foundation for the general theory of heat, for one has not yet succeeded in deriving the laws of thermal equilibrium and the second law of thermodynamics using only the equations of mechanics and the probability calculus, though Maxwell's and Boltzmann's theories came close to this goal.
[1] The purpose of the following considerations is to close this gap. At the same time, they will yield an extension of the second law that is of importance for
[2] the application of thermodynamics. They will also yield the mathematical expression for entropy from the standpoint of mechanics.

§*1. Mechanical model for a physical system*

Let us imagine an arbitrary physical system that can be represented by a mechanical system whose state is uniquely determined by a very large number of
[3] coordinates $p_1 \ldots p_n$ and the corresponding velocities

$$\frac{dp_1}{dt}, \ldots \frac{dp_n}{dt} .$$

Let their energy E consist of two additive terms, the potential energy V and the kinetic energy L. The former shall be a function of the coordinates alone, and the latter shall be a quadratic function of

$$\frac{dp_v}{dt} = p'_\nu ,$$

whose coefficients are arbitrary functions of the p's. Two kinds of external forces shall act upon the masses of the system. One kind of force shall be derivable from a potential V_a and shall represent external conditions (gravity, effect of rigid walls without thermal effects, etc.); their potential may contain time explicitly, but its derivative with respect to time should be very small. The other forces shall not be derivable from a potential and shall vary rapidly. They have to be conceived as the forces that produce the influx of heat. If such forces do not act, but V_a depends explicitly on time, then we are dealing with an adiabatic process.

Also, instead of velocities we will introduce linear functions of them, the momenta $q_1, \ldots, q_n$, as the system's state variables, which are defined by n equations of the form

$$q_\nu = \frac{\partial L}{\partial p'_\nu},$$

where L should be conceived as a function of the $p_1, \ldots, p_n$ and $p'_1, \ldots, p'_n$.

§2. *On the distribution of possible states between N identical adiabatic stationary systems, when the energy contents are almost identical.*

Imagine infinitely many (N) systems of the same kind whose energy content is continuously distributed between definite, very slightly differing values $\bar{E}$ and $\bar{E} + \delta E$. External forces that cannot be derived from a potential shall not be present, and V_a shall not contain the time explicitly, so that the system will be a conservative one. We examine the distribution of states, which we assume to be stationary. [4]

We make the assumption that except for the energy $E = L + V_a + V_i$, or a function of this quantity, for the individual system, there does not exist any function of the state variables p and q which remains constant in time; we shall henceforth consider only systems that satisfy this condition. Our assumption is equivalent to the assumption that the distribution of states of our systems is determined by the value of E and is spontaneously established from any arbitrary initial values of the state variables that satisfy our condition regarding the value of energy. I.e., if there would exist for the [5] [6]

[7] system an additional condition of the kind $\varphi(p_1,\ldots,q_n)$ = const. that cannot be reduced to the form $\varphi(E)$ = const., then it would obviously be possible to choose initial conditions such that each of the N systems could have an arbitrarily prescribed value for φ. However, since these values do not vary with time, it follows, e.g., that for a given value of E any arbitrary value might be assigned to $\Sigma\varphi$, extended over all systems, through appropriate selection of initial conditions. On the other hand, $\Sigma\varphi$ is uniquely calculable by the distribution of states, so that other distributions of states correspond to other values of $\Sigma\varphi$. It is thus clear that the existence of a second such integral φ would necessarily have the consequence that the state distribution would not be determined by E alone but would necessarily have to depend on the initial state of the systems.

If g denotes an infinitesimally small region of all state variables $p_1,\ldots p_n$, $q_1,\ldots q_n$, which is chosen such that $E(p_1\ldots q_n)$ lies between $\bar{E}$ and $\bar{E}+\delta E$ when the state variables belong to the region g, then the distribution of states is characterized by an equation of the form

[8] $$dN = \psi(p_1,\ldots,q_n)\int_g dp_1\ldots dq_n,$$

where dN denotes the number of systems whose state variables belong to the region g at a given time. The equation expresses the condition that the distribution is stationary.

We now choose such an infinitesimal region G. The number of systems whose state variables belong to the region G at a given time $t = 0$ is then

$$dN = \psi(P_1,\ldots Q_n)\int_G dP_1\ldots dQ_n,$$

where the capital letters indicate that the dependent variables pertain to time $t = 0$.

We now let elapse some arbitrary time t. If the system possessed the specific state variables $P_1,\ldots Q_n$ at time $t = 0$, then it will possess the specific state variables $p_1,\ldots,q_n$ at time $t = t$. Systems whose state variables belonged to the region G at $t = 0$, and these systems only, will belong to a specific region g at time $t = t$, so that the following equation applies

$$dN = \psi(p_1, \ldots q_n) \int_g \cdot$$ [9]

However, for each such system Liouville's theorem holds, which has the form

$$\int dP_1, \ldots dQ_n = \int dp_1, \ldots dq_n.$$

From the last three equations it follows that

$$\psi(P_1, \ldots Q_n) = \psi(p_1, \ldots q_n) .$$[1]

Thus, ψ is an invariant of the system, which from the above must have the form $\psi(p_1, \ldots q_n) = \psi^*(E)$. However, for all systems considered, $\psi^*(E)$ differs only infinitesimally from $\psi^*(\bar{E}) =$ const., and our equation of state will then simply be

$$dN = A \int_g dp_1, \ldots dq_n,$$

where A is a quantity independent of the p's and q's.

§3. *On the (stationary) probability of the states of a system S that is mechanically linked with a system Σ whose energy is relatively infinite*

We again consider an infinite number (N) of mechanical systems whose energy shall lie between two infinitesimally different limits $\bar{E}$ and $\bar{E} + \delta\bar{E}$. Let each such mechanical system be, again, a mechanical link between a system S with state variables $p_1, \ldots q_n$ and a system Σ with state variables [11] $\pi_1, \ldots \chi_n$. The expression for the total energy of both systems shall be constituted such that those terms of the energy that accrue through

---

[1] Cf. L. Boltzmann, *Gastheorie* [Theory of gases], Part 2, §32 and §37. [10]

action of the masses of one partial system on the masses of the other partial system are negligible in comparison with the energy E of the partial system [12] S. Further, the energy H of the partial system Σ shall be infinitely large compared with E. Up to the infinitesimally small of higher order, one might then put

$$\mathrm{E} = H + E \, .$$

We now choose a region g that is infinitesimally small in all state variables $p_1 \ldots q_n$, $\pi_1 \ldots \chi_n$ and is so constituted that E lies between the constant values $\bar{\mathrm{E}}$ and $\bar{\mathrm{E}} + \delta\bar{\mathrm{E}}$. The number dN of systems whose state variables belong to the region g is then according to the results of the preceding section

$$dN = A \int_g dp_1 \ldots d\chi_n \, .$$

We note now that we are free to replace A with any continuous function of the energy that assumes the value A for $\mathrm{E} = \bar{\mathrm{E}}$, as this will only infinitesimally change our result. For this function we choose $A' . e^{-2h\mathrm{E}}$, where h denotes a constant which is arbitrary for the time being, and which we will specify soon. We write, then,

[13] $$dN = A' \int_g e^{-2h\mathrm{E}} dp_1 \ldots d\chi_n \, .$$

We now ask: How many systems are in states in which p_1 is between $p_1 + dp_1$, and, respectively, p_2 between $p_2 + dp_2 \ldots$ q_n between q_n and $q_n + dq_n$, but $\pi_1 \ldots \chi_n$ have arbitrary values compatible with the conditions of our system? If we call this number dN', we obtain

$$dN' = A' e^{-2hE} dp_1 \ldots dq_n \int e^{-2hH} d\pi_1 \ldots d\chi_n \, .$$

The integration extends over those values of the state variables for which H lies between $\bar{\mathrm{E}} - E$ and $\bar{\mathrm{E}} - E + \delta\bar{\mathrm{E}}$. We now claim that the value of h can be chosen in one and only one such way that the integral in our equation becomes independent of E.

It is obvious that the integral $\int e^{-2hH} d\pi_1 \ldots d\chi_n$, for which the limits of integration may be determined by the limits E and $\mathrm{E} + \bar{\delta}\mathrm{E}$, will for a [14] specific $\delta\bar{\mathrm{E}}$ be a function of E alone; let us call the latter $\chi(\mathrm{E})$. The integral in the expression for dN' can then be written in the form

$$\chi(\bar{\mathrm{E}} - E) \ .$$

Since E is infinitesimally small compared with $\bar{\mathrm{E}}$, this can be written, up to quantities which are infinitesimally small of higher order, in the form

$$\chi(\bar{\mathrm{E}} - E) = \chi(\bar{\mathrm{E}}) - E\chi'(\bar{\mathrm{E}}) \ .$$

The necessary and sufficient condition for this integral to be independent of E is hence

$$\chi'(\bar{\mathrm{E}}) = 0 \ .$$

But then we can put

$$\chi(\mathrm{E}) = \mathrm{e}^{-2h\mathrm{E}} . \omega(\mathrm{E}) \ ,$$

where $\omega(\mathrm{E}) = \int d\pi_1 \ldots d\chi_n$, extended over all values of the variables whose energy function lies between E and $\mathrm{E} + \delta\mathrm{E}$.

Hence the condition found for h assumes the form

$$e^{-2h\bar{\mathrm{E}}} . \omega(\bar{\mathrm{E}}) . \left\{-2h + \frac{\omega'(\bar{\mathrm{E}})}{\omega(\bar{\mathrm{E}})}\right\} = 0 \ ,$$

or

$$h = \tfrac{1}{2} \frac{\omega'(\bar{\mathrm{E}})}{\omega(\bar{\mathrm{E}})} \ .$$

Thus, there always exists one and only one value for h that satisfies the conditions found. Further, since $\omega(\mathrm{E})$ and $\omega'(\mathrm{E})$ are always positive, as shall be shown in the next section, h is also always a positive quantity.

If we choose h in this way, the integral reduces to a quantity independent of E, so that we obtain the following expression for the number of systems whose variables $p_1 \ldots q_n$ lie within the indicated limits:

[15] $$dN' = A'' e^{-2hE} . dp_1 \ldots dq_n \ .$$

Thus, also for a different meaning of A'', this is the expression for the probability that the state variables of a system mechanically linked with a system of relatively infinite energy lie between infinitesimally close limits when the state has become stationary. [16]

§4. *Proof that the quantity h is positive*

Let $\varphi(x)$ be a homogeneous quadratic function of the variables $x_1 \ldots x_n$. We consider the quantity $z = \int dx_1 \ldots dx_n$, where the limits of integration shall be determined by the condition that $\varphi(x)$ lies between a certain value y and $y + \Delta$, where Δ is a constant. We assert that z, which is a function of y only, always increases with increasing y when $n > 2$.

If we introduce the new variables $x_1 = \alpha x_1' \ldots x_n = \alpha x_n'$, where $\alpha = \text{const.}$, then we have

$$z = \alpha^n \int dx_1' \ldots dx_n' \ .$$

Further, we obtain $\varphi(x) = \alpha^2 \varphi(x')$.

Hence, the limits of integration of the integral obtained for $\varphi(x')$ are

$$\frac{y}{\alpha^2} \quad \text{and} \quad \frac{y}{\alpha^2} + \frac{\Delta}{\alpha^2} \ .$$

Further, if we assume that Δ is infinitesimally small, we obtain

$$z = \alpha^{n-2} \int dx_1' \ldots dx_n' \ .$$

[17] Here y' lies between the limits

$$\frac{y}{a^2} \quad \text{and} \quad \frac{y}{a^2} + \Delta \; .$$

The above equation may also be written as

$$z(y) = a^{n-2} z\left[\frac{y}{a^2}\right] \; .$$

Hence, if we choose a to be positive and $n > 2$, we will always have [18]

$$\frac{z(y)}{z\left[\frac{y}{a^2}\right]} > 1 \; ,$$

which is what had to be proved.

We use this result to prove that h is positive.

We had found

$$h = \tfrac{1}{2} \frac{\omega'(E)}{\omega(E)} \; ,$$

where

$$\omega(E) = \int dp_1 \ldots dq_n \; ,$$

and E lies between E and $E + \delta\bar{E}$. By definition, $\omega(E)$ is necessarily positive, hence we have only to show that $\omega'(E)$ too is always positive.

We choose E_1 and E_2 such that $E_2 > E_1$ and prove that $\omega(E_2) > \omega(E_1)$ and resolve $\omega(E_1)$ into infinitely many summands of the form

$$d\left[\omega(E_1)\right] = dp_1 \ldots dp_n \int dq_1 \ldots dq_n \; .$$

In the integral indicated, the p's have definite values, which are such that $V \leqq E_1$. The limits of integration of the integral are characterized by L lying between $E_1 - V$ and $E_1 + \delta\bar{E} - V$.

To each such infinitesimally small summand corresponds a term out of $\omega(E_2)$ of magnitude

$$d[\omega(E_2)] = dp_1 \ldots dp_n \int dq_1 \ldots dq_n ,$$

where the p's and dp's have the same values as in $d[\omega(E_1)]$, but L lies between the limits $E_2 - V$ and $E_2 - V + \delta\bar{E}$.

Thus, according to the proposition just proved,

$$d[\omega(E_2)] > d[\omega(E_1)] \ .$$

Consequently,

$$\sum d[\omega(E_2)] > \sum d[\omega(E_1)] \ ,$$

where Σ has to be extended over all corresponding regions of the p's.

However,

$$\sum d[\omega(E_1)] = \omega(E_1) \ ,$$

if the summation sign extends over all p's, so that

$$V \leqq E_1 \ .$$

Further, we have

[19] $$\sum d[\omega(E_2)] < \omega(E_2) \ ,$$

since the region of the p's, which is determined by the equation

$$V \leqq E_2$$

includes all of the region defined by the equation

$$V \leqq E_1$$

§5. *On the temperature equilibrium*

We now choose a system S of a specific constitution and call it a thermometer. Let it interact mechanically with the system Σ whose energy is relatively infinitely large. If the state of the entire system is stationary, the state of the thermometer will be defined by the equation

$$dW = Ae^{-2hE}dp_1 \ldots dq_n ,$$

where dW is the probability that the values of the state variables of the thermometer lie within the limits indicated. The constants A and h are related by the equation

$$1 = A. \int e^{-2hE}dp_1 \ldots dq_n ,$$

where the integration extends over all possible values of the state variables. The quantity h thus completely determines the state of the thermometer. We call h the temperature function, noting that, according to the aforesaid, each quantity H observable on the system S must be a function of h alone, as long as V_a remains unchanged, which we have assumed. The quantity h, however, depends only on the state of the system Σ (§3), i.e., it does not depend on the way in which Σ is thermally connected with S. From this we immediately obtain the theorem: If a system Σ is connected with two infinitesimally small thermometers S and S', the same value of h obtains for both thermometers. If S and S' are identical systems, then they will also have identical values of the observable quantity H.

We now introduce only identical thermometers S and call H the observable measure of temperature. We thus arrive at the theorem: The measure of temperature H that is observable on S is independent of the way in which Σ is mechanically connected with S; the quantity H determines h, which in turn determines the energy E of the system Σ, and this in turn determines its state according to our assumption.

From what we have proved it follows immediately that if two systems Σ_1 and Σ_2 are mechanically linked, then they cannot form a system that is in a stationary state unless the two thermometers S connected to them have equal measures of temperature or, what amounts to the same, if they themselves have equal temperature functions. Since the state of the systems Σ_1 and Σ_2 is completely defined by the quantities h_1 and h_2 or H_1 and H_2, it follows that the temperature equilibrium can be determined only by the conditions $h_1 = h_2$ or $H_1 = H_2$.

It now only remains to be shown that two systems that have the same temperature function h (or the same measure of temperature H) can be mechanically connected into one single system that has the same temperature function.

Let two mechanical systems Σ_1 and Σ_2 be merged into one system, but in such a way that the energy terms that contain state variables of both systems be infinitesimally small. Let Σ_1 as well as Σ_2 be connected with an infinitesimally small thermometer S. The readings H_1 and H_2 of the latter are certainly identical up to the infinitesimally small because they refer only to different locations within a single stationary state. The same is of course true of the quantities h_1 and h_2. We now imagine that the energy terms common to both systems decrease infinitely slowly toward zero. Thereby the quantities H and h as well as the distributions of state of the two systems change infinitesimally because they are determined by the energy alone. If then the complete mechanical separation of Σ_1 and Σ_2 is carried out, the relations

$$H_1 = H_2, \qquad h_1 = h_2$$

continue to hold all the same, and the distribution of states changes infin-
[20] itesimally. H_1 and h_1, however, will now pertain only to Σ_1, and H_2 and h_2 only to Σ_2. Our process is strictly reversible, as it consists of a sequence of stationary states. We thus obtain the theorem:

Two systems having the same temperature function h can be merged into a single system having the temperature function h such that their distribution of states changes infinitesimally.

Equality of the quantities h is thus the necessary and sufficient condition for the stationary combination (thermal equilibrium) of two systems. From this follows immediately: If the systems Σ_1 and Σ_2, as well as Σ_1 and Σ_3, can be combined in a stationary fashion mechanically (in thermal equilibrium), then so can Σ_2 and Σ_3.

I would like to note here that until now we have made use of the assumption that our systems are mechanical only inasmuch as we applied Liouville's theorem and the energy principle. Probably the basic laws of the theory of heat can be developed for systems that are defined in a much more general way. We will not attempt to do this here, but will rely on the equations of mechanics. We will not deal here with the important question as to how far the train of thought can be separated from the model employed and generalized.

§6. *On the mechanical meaning of the quantity* h[1]

The kinetic energy L of a system is a homogeneous quadratic function of the quantities q. It is always possible to introduce variables r by a linear substitution such that the kinetic energy will appear in the form [21]

$$L = \tfrac{1}{2}(\alpha_1 r_1^2 + \alpha_2 r_2^2 + \ldots + \alpha_n r_n^2)$$

and that

$$\int dq_1 \ldots dq_n = \int dr_1 \ldots dr_n,$$

when the integral is extended over corresponding infinitesimally small regions. The quantities r are called momentoids by Boltzmann. The mean kinetic energy corresponding to one momentoid when the system together with one of much larger energy forms a single system, assumes the form

$$\frac{\int A'' e^{-2h[V+\alpha_1 r_1^2+\alpha_2 r_2^2+\ldots+\alpha_n r_n^2]} \cdot \frac{\alpha_\nu r_\nu^2}{2} \cdot dp_1 \ldots dp_n \cdot dr_1 \ldots dr_n}{\int A'' e^{-2h[V+\alpha_1 r_1^2+\alpha_2 r_2^2+\ldots+\alpha_n r_n^2]} \cdot dp_1 \ldots dp_n dr_1 \ldots dr_n} = \frac{1}{4h} . \qquad [22]$$

Thus, the mean kinetic energy is the same for all momentoids of a system and is equal to

$$\frac{1}{4h} = \frac{L}{n} , \qquad [24]$$

where L denotes the kinetic energy of the system.

---

[1] Cf. L. Boltzmann, *Gastheorie*, Part 2, §§33, 34, 42. [23]

§7. *Ideal gases. Absolute temperature*

The theory we developed contains as a special case Maxwell's distribution of states for ideal gases. I.e., if in §3 we understand by the system S one gas molecule and by Σ the totality of all the others, then the expression for the probability that the values of the variables $p_1 \ldots p_n$ of S lie in a region g that is infinitesimally small with respect to all variables will be

$$dW = Ae^{-2hE} \int_g dp_1 \ldots dq_n \ .$$

One can also immediately realize from the expression for the quantity h found in §4 that, up to the infinitesimally small, the quantity h will be the same for a gas molecule of another type occuring in the system, since the systems Σ determining h are identical for the two molecules up to the infinitesimally small. This establishes the generalized Maxwellian distribution of states for ideal gases. –

Further, it follows immediately that the mean kinetic energy of motion of the center of gravity of a gas molecule occurring in a system S has the value $\frac{3}{4}h$ because it corresponds to three momentoids. The kinetic theory of gases teaches us that this quantity is proportional to the gas pressure at constant volume. If, by definition, this is taken to be proportional to the absolute temperature, one obtains a relationship of the form

$$\frac{1}{4h} = \kappa . T = \tfrac{1}{2} \frac{\omega(\bar{E})}{\omega'(\bar{E})} \ ,$$

[25] where κ denotes a universal constant, and ω the function introduced in §3.

§8. *The second law of the theory of heat as a consequence of the mechanical theory*

We consider a given physical system S as a mechanical system with coordinates $p_1 \ldots p_n$. As state variables of the system we further introduce the quantities

$$\frac{dp_1}{dt} = p_1' \ldots \frac{dp_n}{dt} = p_n' \ .$$

$P_1 \ldots P_n$ shall be the external forces tending to increase the coordinates of the system. V_i shall be the potential energy of the system, L its kinetic energy, which is a homogeneous quadratic function of the p_ν's. For such a system Lagrange's equations of motion assume the form

$$\frac{\partial(V_i - L)}{\partial p_\nu} + \frac{\mathrm{d}}{dt}\left[\frac{\partial L}{\partial p_\nu'}\right] - P_\nu = 0, \quad (\nu = 1, .. \nu = n) \ .$$

The external forces consist of two kinds of forces. The first kind, $P_\nu^{(1)}$, are the forces that represent the conditions of the system and can be derived from a potential that is a function of $p_1 \ldots p_n$ only (adiabatic walls, gravity, etc.):

$$P_\nu^{(1)} = \frac{\partial V_a}{\partial p_\nu} \ .$$ [26]

Since we have to consider processes which consist of states that infinitely approximate stationary states, we have to assume that even though V_a explicitly contains the time, the partial derivatives of the quantities $\partial V_a / \partial p_\nu$ with respect to time are infinitesimally small.

The second kind of forces, $P_\nu^{(2)} = \Pi_\nu$, shall not be derivable from a potential that depends on the p_ν only. The forces Π represent the forces that mediate the influx of heat.

If one puts $V_a + V_i = V$, equations (1) become [27]

$$\Pi_\nu = \frac{\partial(V - L)}{\partial p_\nu} + \frac{d}{dt}\left\{\frac{\partial L}{\partial p_\nu'}\right\} \ .$$

The work supplied to the system by the forces Π_ν during the time dt represents then the amount of heat dQ absorbed during dt by the system S, which we will measure in mechanical units.

[28]
$$dQ = \sum \Pi_\nu dp_\nu = \sum \frac{\partial V}{\partial p_\nu} dp_\nu - \sum \frac{\partial L}{\partial p_\nu} dp_\nu + \sum \frac{dp_\nu}{dt} \frac{d}{dt} \left[\frac{\partial L}{\partial p_\nu}\right] dt \, .$$

However, since

[29]
$$\sum p'_\nu \frac{d}{dt} \left[\frac{\partial L}{\partial p'_\nu}\right] dt = d \sum p'_\nu \frac{\partial L}{\partial p'_\nu} - \sum \frac{\partial L}{\partial p'_\nu} dp'_\nu$$

and, further,

[30]
$$\sum \frac{\partial L}{\partial p'_\nu} p'_\nu = 2L, \qquad \sum \frac{\partial L}{\partial p'_\nu} dp_\nu + \sum \frac{\partial L}{\partial p'_\nu} dp'_\nu = dL \, ,$$

we have

[31]
$$dQ = \sum \frac{\partial V}{\partial p'_\nu} dp_\nu + dL \, .$$

Since, further

[32]
$$T = \frac{1}{4\kappa h} = \frac{L}{n\kappa} \, ,$$

we will have

[33] (1)
$$\frac{dQ}{T} = n\kappa \frac{dL}{L} + 4\kappa h \sum \frac{\partial V}{\partial p_\nu} dp_\nu \, .$$

We will now concern ourselves with the expression

$$\sum \frac{\partial V}{\partial p_\nu} dp_\nu \, .$$

This represents the increase of potential energy in the system that would take place during time dt if V were not explicitly dependent on time. The time element dt shall be chosen so large that the sum indicated above can be replaced by its average value for infinitely many systems S of equal temperature, and at the same time so small that the explicit changes of h and V
[34] with time be infinitesimally small.

Suppose that infinitely many systems S in a stationary state, all of which have identical h and V_a, change to new stationary systems which are characterized by values $h + \delta h$, $V + \delta V$ common to all. Generally, "δ" shall denote the change of a quantity during transition of the system to a new state; the symbol "d" shall no longer denote the change with time but differentials of definite integrals. –

The number of systems whose state variables lie in the infinitesimally small region g before the change is given by the formula

$$dN = Ae^{-2h(V+L)} \int dp_1 \ldots dp_n \; ; \tag*{[35]}$$

here we are free to choose the arbitrary constant in V for each given h and V_a such that A will equal unity. We shall do this to simplify the calculation and shall call this more precisely defined function V^*.

It can easily be seen that the value of the quantity we seek will be

$$(2) \qquad \sum \frac{\partial V^*}{\partial p_n} dp_n = \frac{1}{N} \int \delta\{e^{-2h(V+L)}\}.V^* dp_1 \ldots dq_n \; , \tag*{[36]}$$

where the integration should extend over all values of the variables, because this expression represents the increase of the mean potential energy of the system that would take effect if the distribution of states would change in conformity with δV^* and δh, but V would not change explicitly.

Further, we obtain

$$(3) \qquad \begin{cases} 4\kappa h \sum \frac{\partial V}{\partial p_\nu} dp_\nu = 4\kappa \frac{1}{N} \int \delta\{e^{-2h(V^*+L)}\}.h.V.dp_1 \ldots dq_n \\ \qquad\qquad = 4\kappa\delta[h\bar{V}] - \frac{4\kappa}{N} \int e^{-2h(V^*+L)} \delta[hV] \; dp_1 \ldots dq_n \; . \end{cases}$$

Here and in the following the integrations have to be extended over all possible values of the variables. Further, it should be kept in mind that the number of systems under consideration does not change. This yields the equation

$$\int \delta(e^{-2h(V^*+L)}) dp_1 \ldots dq_n = 0 \; ,$$

or

$$\int e^{-2h(V^*+L)} \delta(hV) dp_1 \ldots dq_n + \delta h \int e^{-2h(V^*+L)} \delta(L) dp_1 \ldots dq_n = 0 \; , \tag*{[37]}$$

or

$$(4) \qquad \frac{4\kappa}{N} \int e^{-2h(V^*+L)} \delta(hV) dp_1 \ldots dq_n + 4\kappa\bar{L}\delta h = 0 \; .$$

$\bar{V}$ and $\bar{L}$ denote the mean values of the potential and kinetic energies of the N systems. Adding (3) and (4), one obtains

$$4\kappa h \sum \frac{\partial V^*}{\partial p_\nu}\, dp_\nu = 4\kappa\delta[h\bar{V}] + 4\kappa\bar{L}.\delta h \ ,$$

or, because

[38] $$h = \frac{n}{4\bar{L}} \ , \qquad \delta h = -\frac{n}{4L^2}\cdot\delta L \ ,$$

$$4\kappa h \sum \frac{\partial V}{\partial p_\nu}\, dp_\nu = 4\kappa\delta[h\bar{V}] - n\kappa\frac{\delta L}{L} \ .$$

If we substitute this formula in (1), we obtain

[39] $$\frac{dQ}{T} = \delta[4\kappa h\bar{V}^*] = \delta\left[\frac{V^*}{T}\right] \ .$$

Thus, dQ/T is a complete differential. Since

$$\frac{\bar{L}}{T} = n\kappa \ , \qquad \text{thus} \quad \delta\left[\frac{L}{T}\right] = 0 \ ,$$

one may also set

$$\frac{dQ}{T} = \delta\left[\frac{E^*}{T}\right] \ .$$

Thus, apart from an arbitrary additive constant, E^*/T is the expression for the entropy of the system, where we have put $E^* = V^* + L$. The second law thus appears as a necessary consequence of the mechanistic world picture.

§9. *Calculation of the entropy*

The expression $\epsilon = E^*/T$ that we obtained for the entropy ϵ only appears to be simple, because E^* remains to be calculated from the conditions of the mechanical system. I.e., we have

$$E^* = E + E_0 \ ,$$

where E is given directly, but E_0 has to be determined as a function of E and h from the condition

$$\int e^{-2h(E-E_0)}dp_1\ldots dq_n = N\ .$$ [40]

In this way, one obtains

$$\epsilon = \frac{E^*}{T} = \frac{E}{T} + 2\kappa \log\left[\int e^{-2hE}dp_1\ldots dq_n\right] + \text{const.}$$ [41]

In the expression thus obtained, the arbitrary constant that has to be added to the quantity E does not affect the result, and the third term, denoted "const.," is independent of V and T.

The expression for the entropy ϵ is strange, because it depends solely on E and T, but no longer reveals the special form of E as the sum of potential and kinetic energy. This fact suggests that our results are more general than the mechanical model used, the more so as the expression for h found in §3 shows the same property. [42]

§*10*. *Extension of the second law*

No assumptions had to be made about the nature of the forces that correspond to the potential V_a, not even that such forces occur in nature. Thus, the mechanical theory of heat requires that we arrive at correct results if we apply Carnot's principle to ideal processes, which can be produced from the observed processes by introducing arbitrarily chosen V_a's. Of course, the results obtained from the theoretical consideration of those processes have a real meaning only when the ideal auxiliary forces V_a no longer appear in them. [43]

Bern, June 1902. (Received on 26 June 1902)

Doc. 4

A THEORY OF THE FOUNDATIONS OF THERMODYNAMICS

by A. Einstein

[*Annalen der Physik* 11 (1903): 170-187]

[1] In a recently published paper I showed that the laws of thermal equilibrium and the concept of entropy can be derived with the help of the kinetic theory of heat. The question that then arises naturally is whether the kinetic theory is really necessary for the derivation of the above foundations of the theory of heat, or whether perhaps assumptions of a more general nature may suffice. In this article it shall be demonstrated that the latter is the [2] case, and it shall be shown by what kind of reasoning one can reach the goal.

§1. On a general mathematical representation of the processes in isolated physical systems

Let the state of some physical system that we consider be uniquely determined by very many (n) scalar quantities $p_1, p_2 \ldots p_n$, which we call [3] state variables. The change of the system in a time element dt is then determined by the changes $dp_1, dp_2 \ldots dp_n$ that the state variables undergo during that time element.

Let the system be isolated, i.e., the system considered should not interact with other systems. It is then clear that the state of the system at a given instant of time uniquely determines the change of the system in the next time element dt, i.e., the quantities $dp_1, dp_2 \ldots dp_n$. This statement is equivalent to a system of equations of the form

$$(1) \qquad \frac{dp_i}{dt} = \varphi_i(p_1 \ldots p_n) \qquad (i = 1 \ \ldots \ i = n) \ ,$$

where the φ's are unique functions of their arguments.

In general, for such a system of linear differential equations there does not exist an integral of the form

$$\psi(p_1 \ldots p_n) = \text{const.},$$

which does not contain the time explicitly. However, for a system of equations that represents the changes of a physical system closed to the outside, we must assume that at least one such equation exists, namely the energy equation

$$E(p_1 \ldots p_n) = \text{const.}$$

At the same time, we assume that no further integral of this kind that is independent of the above equation is present. [4]

§2. *On the stationary distribution of state of infinitely many isolated physical systems of almost equal energies*

Experience shows that after a certain time an isolated system assumes a state in which no perceptible quantity of the system undergoes any further changes with time; we call this state the stationary state. Hence it will obviously be necessary for the functions φ_i to fulfill a certain condition so that equations (1) may represent such a physical system.

If we now assume that a perceptible quantity is always represented by a time average of a certain function of the state variables $p_1 \ldots p_n$, and that these state variables $p_1 \ldots p_n$ keep on assuming the same systems of values with always the same unchanging frequency, then it necessarily follows from this condition, which we shall elevate to a postulate, that the averages of all functions of the quantities $p_1 \ldots p_n$ must be constant; hence, in accordance with the above, all perceptible quantities must also be constant.

We will specify this postulate precisely. Starting at an arbitrary point of time and throughout time T, we consider a physical system that is represented by equations (1) and has the energy E. If we imagine having chosen some arbitrary region Γ of the state variables $p_1 \ldots p_n$, then at a given instant of time T the values of the variables $p_1 \ldots p_n$ will lie within the chosen region Γ or outside it; hence, during a fraction of the time T, which we shall call τ, they will lie in the chosen region Γ. Our condition then reads as follows: If $p_1 \ldots p_n$ are state variables of a physical system, i.e., of a system that assumes a stationary state, then for each region Γ the quantity τ/T has a definite limiting value for $T = \infty$. For any infinitesimally small region this limiting value is infinitesimally small.

The following consideration can be based on this postulate. Let there be very many (N) independent physical systems, all of which are represented by the same system of equations (1). We select an arbitrary instant t and inquire after the distribution of the possible states among these N systems, assuming that the energy E of all systems lies between E^* and the infinitesimally close value $E^* + \delta E^*$. From the postulate introduced above, it follows immediately that the probability that the state variables of a system randomly selected from among N systems will lie within the region Γ at time t has the value

$$\lim_{T=\infty} \frac{\tau}{T} = \text{const.}$$

The number of systems whose state variables lie within the region Γ at time t is thus

$$N \cdot \lim_{T=\infty} \frac{\tau}{T} ,$$

i.e., a quantity independent of time. If g denotes a region of the coordinates $p_1 \ldots p_n$ that is infinitesimally small in all variables, then the number of systems whose state variables fill up an arbitrarily chosen infinitesimally small region g at an arbitrary time will be

[5] (2)
$$dN = \epsilon(p_1 \ldots p_n) \int_g dp_1 \ldots dp_n \ .$$

The function ϵ is obtained by expressing in symbols the condition that the distribution of states expressed by equation (2) is a stationary one. Specifically, the region g shall be chosen such that p_1 shall lie between the definite values p_1 and $p_1 + dp_1$, p_2 between p_2 and $p_2 + dp_2, \ldots p_n$ between p_n and $p_n + dp_n$; then we have at the time t

$$dN_t = \epsilon(p_1 \ldots p_n) . dp_1 . dp_2 \ldots dp_n ,$$

where the subscript of dN denotes the time. Taking into account equation (1), one obtains furthermore at time $t + dt$ and the same region of the state variables

$$dN_{t+dt} = dN_t - \sum_{\nu=1}^{\nu=n} \frac{\partial(\epsilon\varphi_\nu)}{\partial p_\nu} .dp_1 \ldots dp_n .dt \ .$$

However, since $dN_t = dN_{t+dt}$, because the distribution is stationary, we have

$$\sum \frac{\partial(\epsilon\varphi_\nu)}{\partial p_\nu} = 0 \ . \tag*{[6]}$$

This yields

$$-\sum \frac{\partial\varphi_\nu}{\partial p_\nu} = \sum \frac{\partial(\log\,\epsilon)}{\partial p_\nu} \cdot \varphi_\nu = \sum \frac{\partial(\log\,\epsilon)}{\partial p_\nu} \cdot \frac{dp_\nu}{dt} = \frac{d(\log\,\epsilon)}{dt} \ ,$$

where $d(\log\,\epsilon)/dt$ denotes the change of the function $\log\,\epsilon$ with respect to time for an individual system, taking into account the changes with time of the quantities p_ν.

One obtains further

$$\epsilon = e^{-\int dt \sum_{\nu=1}^{\nu=n} \frac{\partial\varphi_\nu}{\partial p_\nu} + \psi(E)} = e^{-m+\psi(E)} \ .$$

The unknown function ψ is the time-independent integration constant which may depend on the variables $p_1 \ldots p_n$, but can contain them, according to the assumptions made in §1, only in the combination in which they appear in the energy E.

However, since $\psi(E) = \psi(E^*) = \text{const.}$ for all N systems considered, the expression for ϵ reduces in our case to

$$\epsilon = \text{const.}\ e^{-\int dt \sum_{\nu=1}^{\nu=n} \frac{\partial\varphi_\nu}{\partial p_\nu}} = \text{const.}\ e^{-m} \ .$$

According to the above we now have

$$dN = \text{const.}\ e^{-m} \int_g dp_1 \ldots dp_n \ .$$

For the sake of simplicity we now introduce new state variables for the system considered; they shall be denoted by π_ν. We then have

$$dN = \frac{e^{-m}}{\frac{D(\pi_1 \ldots \pi_n)}{D(p_1 \ldots p_n)}} \int_g d\pi_1 \ldots d\pi_n \ ,$$

where the symbol D denotes the functional determinant. – We now want to choose the new coordinates such that

$$e^{-m} = \frac{D(\pi_1 \ldots \pi_n)}{D(p_1 \ldots p_n)} \ .$$

This equation can be satisfied in infinitely many ways, e.g., by setting

$$\begin{array}{l} \pi_2 = p_2 \\ \pi_3 = p_3 \\ \ldots\ldots\ldots \\ \pi_n = p_n \end{array} \qquad \pi_1 = \int e^{-m} \cdot dp_1 \ .$$

Using the new variables, we thus obtain

$$dN = \text{const.} \int d\pi_1 \ldots d\pi_n \ .$$

[7] Henceforth we will always suppose that such variables have been introduced.

§3. *On the distribution of state of a system in contact with a system of relatively infinitely large energy.*

We now assume that each of the N isolated systems is composed of two partial systems Σ and σ in interaction. Let the state of the partial system Σ be determined by the values of the variables $\Pi_1 \ldots \Pi_\lambda$, and that of the system σ by the values of the variables $\pi_1 \ldots \pi_\ell$. Further, let the energy E, which for each system shall lie between the values E^* and δE^*, i.e., shall equal E^* up to the infinitesimally small, be composed of two terms, of which the first, H, shall be determined only by the values of the state variables of Σ, and the second, η, only by the state variables of σ, so that, except for the relatively infinitesimally small, one has

$$E = H + \eta \ .$$ [8]

Two systems in interaction which satisfy this condition will be called two systems in contact. We also assume that η is infinitesimally small compared with H.

For the number dN_1 of the N-systems whose state variables $\Pi_1 \ldots \Pi_\lambda$ and $\pi_1 \ldots \pi_\ell$ lie between Π_1 and $\Pi_1 + d\Pi_1$, Π_2 and $\Pi_2 + d\Pi_2$, $\ldots \Pi_\lambda$ and $\Pi_\lambda + d\Pi_\lambda$, and π_1 and $\pi_1 + d\pi_1$, π_2 and $\pi_2 + d\pi_2$, $\ldots \pi_\ell$ and $\pi_\ell + d\pi_\ell$, we get the expression

$$dN_1 = C \cdot d\Pi_1 \ldots d\Pi_\lambda \cdot d\pi_1 \ldots d\pi_\ell \ ,$$

where C can be a function of $E = H + \eta$.

However, since according to the above assumption the energy of each of the systems considered up to the infinitesimally small has the value E^*, we can replace C by $\text{const.}\, e^{-2hE^*} = \text{const.}\, e^{-2h(H+\eta)}$ without causing any changes in the result, where h is a constant still to be defined precisely. Hence, the expression for dN_1 becomes

$$dN_1 = \text{const.}\ e^{-2h(H+\eta)} \cdot d\Pi_1 \ldots d\Pi_\lambda \cdot d\pi_1 \ldots d\pi_\ell \ .$$

The number of systems whose state variables π lie between the indicated limits, while the values of the variables Π are not subjected to any restrictive condition, may thus be represented in the form

$$dN_2 = \text{const.}\ e^{-2h\eta} \cdot d\pi_1 \ldots d\pi_\ell \int e^{-2hH} d\Pi_1 \ldots d\Pi_\lambda \ ,$$

where the integral is to be extended over all values of Π to which correspond values of the energy H lying between $E^* - \eta$ and $E^* + \delta E^* - \eta$. Had the integration been carried out, we would have found the distribution of the state of the systems σ. This is in fact possible.

We put

$$\int e^{-2hH} \cdot d\Pi_1 \ldots d\Pi_\lambda = \chi(E) ,$$

where the integral on the left-hand side is to be extended over all values of the variables for which H lies between the definite values E and $E + \delta E^*$. The integral that appears in the expression dN_2 then assumes the form

$$\chi(E^* - \eta) ,$$

or, since η is infinitesimally small compared with E^*,

$$\chi(E^*) - \chi'(E^*) \cdot \eta .$$

Thus, if h can be chosen such that $\chi'(E^*) = 0$, the integral reduces to a quantity that is independent of the state of σ.

It is possible to put, up to the infinitesimally small,

$$\chi(E) = e^{-2hE} \int d\Pi_1 \ldots d\Pi_\lambda = e^{-2hE} \cdot \omega(E) ,$$

[9] where the integration limits are the same as above, and where ω denotes a new function of E.

The condition for h now assumes the form

$$\chi'(E^*) = e^{-2hE^*} \cdot \{\omega'(E^*) - 2h\omega(E^*)\} = 0 ;$$

consequently:

$$h = \frac{1}{2} \frac{\omega'(E^*)}{\omega(E^*)} .$$

If h is chosen in this way, the expression for dN_2 will assume the form

(3) $$dN_2 = \text{const. } e^{-2h\eta} d\pi_1 \ldots d\pi_\ell \ .$$

With suitable choice of the constant this expression represents the probability that the state variables of a system in contact with another system of relatively infinitely large energy will lie within the indicated limits. The quantity h depends only on the state of the above system Σ of relatively infinitely large energy. [10]

§4. *On absolute temperature and thermal equilibrium*

Thus, the state of the system σ depends only on the quantity h, and the latter only on the state of the system Σ. We call the quantity $1/4h\kappa = T$ the absolute temperature of the system Σ, where κ denotes a universal constant. [11] [12]

If we call the system σ "thermometer," then we can immediately advance the following propositions:

1. The state of the thermometer depends only on the absolute temperature of the system Σ, and not on the kind of contact of the systems Σ and σ.

2. If in case of contact two systems Σ_1 and Σ_2 impart the same state to a thermometer σ, then they have the same absolute temperature and will also impart the same state to another thermometer σ' in case of contact.

Further, suppose two systems Σ_1 and Σ_2 are in contact and Σ_1 is also in contact with a thermometer σ. The distribution of states of σ depends then only on the energy of the system $(\Sigma_1 + \Sigma_2)$, i.e., on the quantity $h_{1,2}$. If the interaction between Σ_1 and Σ_2 is imagined to decrease infinitely slowly, this does not change the expression for the energy $H_{1,2}$ of the system $(\Sigma_1 + \Sigma_2)$, which can be readily seen from our definition of contact and the expression for the quantity h that we formulated in the last section. Finally, if the interaction had ceased completely, the distribution of states of σ, which does not change during the separation of Σ_1 and Σ_2, will now depend on Σ_1, i.e., on the quantity h_1, where the index denotes association with the system Σ_1 alone. Hence we have

$$h_1 = h_{12} .$$

By an analogous line of argument, one could have obtained

$$h_2 = h_{12} ;$$

hence

$$h_1 = h_2 ,$$

or, in words: If one separates two systems Σ_1 and Σ_2 in contact which form an isolated system $(\Sigma_1 + \Sigma_2)$ of absolute temperature T, then the now [13] isolated systems Σ_1 and Σ_2 will have the same temperature after separation. We imagine a given system in contact with an ideal gas. This gas shall be completely describable in terms of the kinetic theory of gases. As the system σ we consider a single monoatomic gas molecule of mass μ whose state shall be completely determined by its orthogonal coordinates x, y, z and the velocities ξ, η, ζ. In accordance with §3, we obtain for the probability that the state variables of this molecule lie between the limits x and $x + dx$, ...ζ and $\zeta + d\zeta$ the well-known Maxwellian expression

$$dW = \text{const.}\ e^{-h\mu(\xi^2+\eta^2+\zeta^2)} \cdot dx \ldots d\zeta .$$

By integration, one obtains from this for the mean kinetic energy of this molecule

[14] $$\overline{\frac{\mu}{2}(\xi^2 + \eta^2 + \zeta^2)} = \frac{1}{4h} .$$

However, the kinetic theory of gases teaches that at constant volume of the gas this quantity is proportional to the pressure exerted by the gas. The latter is by definition proportional to the quantity designated in physics as absolute temperature. Thus the quantity we designated as absolute temperature is nothing else but the temperature of a system measured by the gas thermometer.

§5. *On infinitely slow processes*

Until now we have only considered systems that are in a stationary state. Now we are also going to investigate changes of stationary states, though only those that proceed so slowly that the distribution of states existing at an arbitrary instant differs only infinitesimally from the stationary distribution; or, more precisely, that, up to the infinitesimally small, the probability that the state variables lie in a certain region G can be represented at any moment by the formula found above. We call such a change an infinitesimally slow process.

If the functions φ_ν (equation (1)) and the energy E of a system are specified, then, according to the above, its stationary state distribution is also specified. An infinitely slow process will thus be specified either by a changing E, or by the functions φ_ν containing the time explicitly, or by both circumstances simultaneously, but in such a way that the corresponding differential quotients with respect to time are very small.

We assumed that the state variables of an isolated system change according to equations (1). However, conversely, if there exists a system of equations (1) according to which the state variables of a system are changing, this system does not always have to be an isolated one. For it can happen that a system under consideration is influenced by other systems in such a way that this influence depends only on such functions of the variable coordinates of the influencing systems which do not change when the distribution of states of the influencing system is constant. In this case the change of the coordinates p_ν of the system considered can also be represented by a system having the form of equations (1). However, the functions φ_ν will then depend not only on the physical nature of the system in question, but also on certain constants that are defined through the influencing systems and their distributions of states. This kind of influence on the system under consideration we call adiabatic. It is easy to see that as long as the distributions of state of the adiabatically influencing systems do not change, there exists an energy equation for the equations (1) in this case as well. If the states of the adiabatically influencing systems do change, then the functions φ_ν of the [15] systems considered change explicitly with time, with equations (1) maintaining their validity at all times. Such a change of the distribution of states of the system under consideration we call an adiabatic one.

We now consider a second kind of changes of the state of a system Σ. Consider a system that can be influenced adiabatically. We assume that at time $t = 0$ the system Σ enters into such an interaction with a system P of a different temperature that we called "in contact" above, and we remove the system P after the time necessary for the equalization of the temperatures of Σ and P. The energy of Σ has then changed. The equations (1) of Σ are invalid during the process but valid before and after it, while the functions φ_ν are the same before and after the process. Such [16] a process we call "isopycnic" and the energy supplied to Σ, "heat supplied."

It is evident that, up to the infinitesimally small, it is possible to construct each infinitely slow process from a succession of infinitesimally small adiabatic and isopycnic processes, so that in order to get a general overview we have to study the latter ones only.

§6. *On the concept of entropy*

Let there be a physical system whose instantaneous state shall be completely determined by the values of the state variables $p_1 \ldots p_n$. Let this system undergo a small, infinitely slow process, in which the systems that influence this system adiabatically experience an infinitesimally small change of state, and energy is being supplied to the system considered by systems in contact. We take account of the adiabatically influencing systems by stipulating that in addition to the $p_1 \ldots p_n$, the energy E of the system considered shall also depend on some parameters $\lambda_1, \lambda_2 \ldots$, whose values shall be determined by the distributions of states of the systems that influence adiabatically the system considered. In purely adiabatic processes there holds at any instant a system of equations (1) whose functions φ_ν depend not only on the coordinates p_ν but also on the slowly changing quantities λ; for adiabatic processes too, there will hold at any instant the energy equation, whose form is

$$\sum \frac{\partial E}{\partial p_\nu} \varphi_\nu = 0 \ .$$

We now investigate the energy increase of the system during an arbitrary infinitesimally small, infinitely slow process.

For each time element dt of the process we have

$$dE = \sum \frac{\partial E}{\partial \lambda} d\lambda + \sum \frac{\partial E}{\partial p_\nu} dp_\nu \ . \tag{4}$$

For an infinitesimally small isopycnic process, all $d\lambda$ vanish in each time element, and thus the first term of the right-hand side vanishes too. However, since according to the previous section, in an isopycnic process dE is to be considered as heat supplied, for such a process the heat supplied dQ is represented by the expression

$$dQ = \sum \frac{\partial E}{\partial p_\nu} dp_\nu \ .$$

However, for an adiabatic process, during which equations (1) are always valid, we have, according to the energy equation,

$$\sum \frac{\partial E}{\partial p_\nu} dp_\nu = \sum \frac{\partial E}{\partial p_\nu} \varphi_\nu dt = 0 \ .$$

On the other hand, according to the previous section, $dQ = 0$ for an adiabatic process, so that one can put

$$dQ = \sum \frac{\partial E}{\partial p_\nu} dp_\nu$$

for an adiabatic process as well. Hence, this equation must be considered as valid for any arbitrary process during each time element. Thus equation (4) becomes

$$dE = \sum \frac{\partial E}{\partial \lambda} d\lambda + dQ \ . \tag{4'}$$

This expression represents the energy change of the system occurring during the whole infinitesimally small process at changed values of $d\lambda$ and dQ as well.

At the beginning and the end of the process, the distribution of states of the system considered is stationary, and when the system is in contact with a system of relatively infinitely large energy before and after the process, this assumption having formal significance only, this distribution is defined by the equation having the form

$$\begin{aligned} dW &= \text{const.}\ e^{-2hE}\cdot dp_1\ldots dp_n \\ &= e^{c-2hE}\cdot dp_1\ldots dp_n\ , \end{aligned}$$

where dW denotes the probability that the values of the system's state variables lie within the limits indicated at any arbitrarily chosen moment. The constant c is defined by the equation

$$\int e^{c-2hE}\cdot dp_1\ldots dp_n = 1\ , \tag{5}$$

where the integration has to be extended over all values of the variables.

Specifically, if equation (5) holds before the process under consideration, then afterwards we have

$$\int e^{(c+dc)-2(h+dh)\left[E\,+\,\sum\frac{\partial E}{\partial\lambda}d\lambda\right]}\cdot dp_1\ldots dp_n = 1\ , \tag{5'}$$

and the two last equations yield

$$\int\left[dc\ -\ 2Edh\ -\ 2h\sum\frac{\partial E}{\partial\lambda}\cdot d\lambda\right]\cdot e^{c-2hE}\cdot dp_1\ldots dp_n = 0\ ,$$

or, since the expression in parentheses can be taken as a constant during integration because the system's energy E never differs markedly from a fixed average value before and after the process, and taking into account equation (5),

$$dc\ -\ 2Edh\ -\ 2h\sum\frac{\partial E}{\partial\lambda}\,d\lambda = 0\ . \tag{5''}$$

However, according to equation (4') we have

$$-2hdE + 2h\sum \frac{\partial E}{\partial \lambda}\, d\lambda + 2hdQ = 0 \ ,$$

and by adding these two equations one obtains

$$2h\cdot dQ = d(2hE - c) \ ,$$

or, since $1/4h = \kappa.T$,

$$\frac{dQ}{T} = d\left[\frac{E}{T} - 2\kappa c\right] = dS \ .$$

This equation states that dQ/T is a total differential of a quantity that we will call the entropy S of the system. Taking into account equation (5), one obtains

$$S = 2\kappa(2hE - c) = \frac{E}{T} + 2\kappa \log \int e^{-2hE} dp_1 \ldots dp_n \ ,$$

where the integration has to be extended over all values of the variables.

§7. *On the probability of distributions of states*

In order to derive the second law in its most general form, we have to investigate the probability of distributions of states.

We consider a very large number (N) of isolated systems, all of which can be represented by the same system of equations (1), and whose energies coincide up to the infinitesimally small. The distribution of states of these N systems can then be represented by an equation of the form

$$(2') \qquad dN = \epsilon(p_1 \ldots p_n, t)\, dp_1 \ldots dp_n \ ,$$

where in general ϵ depends explicitly on the state variables $p_1 \ldots p_n$ and also on time. Here the function ϵ completely characterizes the distribution of states.

It follows from §2 that when the distribution of states is constant, which, according to our assumptions, is always the case at very large values of t, we must have ϵ = const., so that for a stationary distribution of states we will have

$$dN = \text{const.}\ dp_1 \ldots dp_n \ .$$

From this it follows immediately that the expression for the probability dW for the values of the state variables of a system randomly chosen from among the N systems to lie in the infinitesimally small region g of the state variables located within the assumed energy limits is given by

$$dW = \text{const.} \int_g dp_1 \ldots dp_n \ .$$

This proposition can also be formulated as follows: If the whole pertinent region of state variables that is determined by the assumed energy limits is divided into ℓ partial regions $g_1, g_2 \ldots g_\ell$ such that

$$\int_{g_1} = \int_{g_2} = \cdots = \int_{g_\ell} \ ,$$

and if one denotes by W_1, W_2, etc., the probabilities that the values of the state variables of the arbitrarily chosen system lie within $g_1, g_2 \ldots$ at a certain instant, then

$$W_1 = W_2 = \cdots W_\ell = \frac{1}{\ell} \ .$$

The probability that at a given moment the system considered will belong to a specific region from among these $g_1 \ldots g_\ell$ regions is thus just as great as the probability that it will belong to any other of these regions.

The probability that, at a randomly chosen time, ϵ_1 of N systems considered will belong to the region g_1, ϵ_2 to region g_2, $\ldots \epsilon_\ell$ to region g_ℓ, is hence

$$W = \left[\frac{1}{\ell}\right]^N \frac{N!}{\epsilon_1!\epsilon_2!\ldots\epsilon_n!} \ ,$$

or also, since $\epsilon_1, \epsilon_2 \ldots \epsilon_n$ are to be thought of as very large numbers:

$$\log W = \text{const.} - \sum_{\epsilon=1}^{\epsilon=\ell} \epsilon \log \epsilon \ .$$

If ℓ is sufficiently large, one can put without noticeable error

$$\log W = \text{const.} - \int \epsilon \log \epsilon \, dp_1 \ldots dp_n \ .$$

In this equation W denotes the probability that a given distribution of states, which is expressed by the numbers $\epsilon_1, \epsilon_2 \ldots \epsilon_\ell$, or, else, by a specific function ϵ of $p_1 \ldots p_n$ according to equation (2'), prevails at a given time.

If in this equation ϵ were constant, i.e., independent of the p_ν's within the energy limits considered, then the distribution of states considered would be stationary, and, as can easily be proved, the expression for the probability W of the distribution of states would be a maximum. If ϵ depends on the values of the p_ν's, then it can be shown that the expression for $\log W$ for the distribution of states considered does not have an extremum, i.e., that there exist distributions of states differing infinitesimally from the considered one for which W is larger.

If we follow the N systems considered for an arbitrary time interval, the distribution of states, and thus also W, will continually change with time, and we will have to assume that always more probable distributions of states will follow upon improbable ones, i.e., that W increases until the distribution of states has become constant and W a maximum. [17]

It will be shown in the following sections that the second law of thermodynamics can be deduced from this proposition.

First of all, we have

$$- \int \epsilon' \log \epsilon' \, dp_1 \ldots dp_n \geqq - \int \epsilon \log \epsilon \, dp_1 \ldots dp_n \ ,$$

where the function ϵ determines the distribution of states of the N systems at a certain time t, the function ϵ' determines the distribution of states at a certain later time t', and the integration on both sides is to be extended over all values of the variables. Further, if the quantities $\log \epsilon$ and $\log \epsilon'$ of the individual systems from among the N systems do not differ markedly from each other, then, since

$$\int \epsilon dp_1 \ldots dp_n = \int \epsilon' dp_1 \ldots dp_n = N ,$$

the last equation becomes

$$- \log \epsilon' \geqq - \log \epsilon . \tag{6}$$

§8. *Application of the results obtained to a particular case*

We consider a finite number of physical systems $\sigma_1, \sigma_2 \ldots$ that together form an isolated system, which we shall call total system. The systems $\sigma_1, \sigma_2 \ldots$ shall not interact markedly with each other thermally, but they might affect each other adiabatically. The distribution of states of each of the systems $\sigma_1, \sigma_2 \ldots$, which we shall call partial systems, shall be stationary up to the infinitesimally small. The absolute temperatures of the partial systems may be arbitrary and different from each other.

The distribution of states of the system σ_1 will not be markedly different from the distribution of states that would hold if σ_1 were in contact with a physical system of the same temperature. We can therefore represent its distribution of states by the equation

$$dw_1 = e^{c_{(1)} - 2h_{(1)}E_{(1)}} \int_g dp_1^{(1)} \ldots dp_{(n)}^{(1)},$$

where the indices (1) indicate affiliation with the partial system σ_1.

Analogous equations hold for the other partial systems. Since the instantaneous values of the state variables of the individual partial systems are independent of those of the other systems, we obtain for the distribution of states of the total system an equation of the form

[18] (7) $$dw = dw_1 \cdot dw_2 \cdots = \mathrm{e}^{\sum c_\nu - 2h_\nu E_\nu} \int_g dp_1 \ldots dp_n ,$$

where the summation is to be extended over all systems, and the integration over the arbitrary region g, which is infinitesimally small in all the variables of the total system.

We now assume that after some time the partial systems $\sigma_1, \sigma_2 \dots$ enter into some arbitrary interaction with each other, but that during that process the total system always remains an isolated one. After the lapse of a certain time there shall arise a state of the total system in which the partial systems $\sigma_1, \sigma_2 \dots$ do not affect each other thermally and, up to the infinitesimally small, exist in a stationary state.

Then an equation completely analogous to that holding before the process will hold for the distribution of states of the total system:

$$(7') \qquad dw' = dw_1' \cdot dw_2' \cdots = e^{\sum(c_\nu' - 2h_\nu' E_\nu')} \int_g dp \dots dp_n \, .$$

We now consider N such total systems. Up to the infinitesimally small, equation (7) shall hold for each of these systems at time t, and equation (7') at time t'. Then the distribution of states of the N total systems considered at times t and t' will be given by the equations

$$dN_t = N \cdot e^{\sum(c_\nu - 2h_\nu E_\nu)} \cdot dp_1 \dots dp_n$$
$$dN_{t'} = N \cdot e^{\sum(c_\nu' - 2h_\nu' E_\nu')} \cdot dp_1 \dots dp_n \, .$$

To these two distributions of states we now apply the results of the previous section. Neither the

$$\epsilon = N \cdot e^{\sum(c_\nu - 2h_\nu E_\nu)}$$

nor the

$$\epsilon' = N \cdot e^{\sum(c_\nu' - 2h_\nu' E_\nu')}$$

for the individual systems among the N systems are here markedly different, so that we can apply equation (6), which yields

$$\sum (2h'E' - c') \geqq \sum (2hE - c) \, ,$$

or, noting that according to §6 the quantities $2h_1E_1 - c_1$, $2h_2E_2 - c_2, \dots$ are identical with the entropies $S_1, S_2 \dots$ of the partial systems up to a universal constant,

$$(8) \qquad S'_1 + S'_2 + \cdots \geqq S_1 + S_2 + \cdots ,$$

i.e., the sum of the entropies of the partial systems of an isolated system after some arbitrary process is equal to or larger than the sum of the entropies of the partial systems before the process.

§9. *Derivation of the second law*

Let there be an isolated total system whose partial systems shall be called W, M, and $\Sigma_1, \Sigma_2 \dots$. Let the system W, which we shall call heat reservoir, have an energy that is infinitely large compared with the system M (engine). Similarly, the energy of the systems $\Sigma_1, \Sigma_2 \dots$, which interact adiabatically with each other, shall be infinitely large compared with that of M. We assume that all the partial systems M, W, Σ_1, $\Sigma_2 \dots$ are in a stationary state.

Suppose that the engine M passes through a cyclic process during which it changes the distributions of states of the systems $\Sigma_1, \Sigma_2 \dots$ infinitely slowly through adiabatic influence, i.e., performs work, and receives the amount of heat Q from the system W. The reciprocal adiabatic influence of the systems $\Sigma_1, \Sigma_2 \dots$ at the end of the process will then differ from that before the process. We say that the engine M has converted the amount of heat Q into work.

We now calculate the increase in entropy of the individual partial systems during the process considered. According to the results of §6 the entropy increase of the heat reservoir W equals $-Q/T$ if T denotes the absolute temperature. The entropy of M is the same before and after the process because the system M has undergone a cyclic process. The systems $\Sigma_1, \Sigma_2 \dots$ do not change their entropies during the process at all because these systems only experience an adiabatic influence that is infinitely slow. Hence the entropy increase $S' - S$ of the total system has the value

$$S' - S = -\frac{Q}{T} .$$

Since according to the results of the last section this quantity $S' - S$ is always $\geqq 0$, it follows that

$$Q \leqq 0 .$$

This equation expresses the impossibility of the existence of a perpetuum mobile of the second kind.

Bern, January 1903. (Received on 26 January 1903)

Doc. 5

ON THE GENERAL MOLECULAR THEORY OF HEAT

by A. Einstein

[*Annalen der Physik* 14 (1904): 354-362]

In the following I present a few addenda to an article I published last year.[1]

When I refer to the "general molecular theory of heat," I mean a theory that is essentially based on the assumptions put forth in §1 of the article cited. In order to avoid unnecessary repetitions, I assume that the reader is familiar with that article and use the same notations I have used there.

First, I derive an expression for the entropy of a system, which is
[2] completely analogous to the expression found by Boltzmann for ideal gases and
[3] assumed by Planck in his theory of radiation. Then I give a simple derivation of the second law. After that I examine the meaning of a universal constant,
[4] which plays an important role in the general molecular theory of heat. I conclude with an application of the theory to black-body radiation, which yields a most interesting relationship between the above-mentioned universal
[5] constant, which is determined by the magnitudes of the elementary quanta of matter and electricity, and the order of magnitude of the radiation wavelengths, without recourse to special hypotheses.

§*1*. *On the expression for entropy*

For a system that can absorb energy only in the form of heat, or, in other words, for a system not affected adiabatically by other systems, the following equation holds between the absolute temperature T and the energy E, according to §3 and §4, loc.cit.:

---

[1] [1] A. Einstein, *Ann. d. Phys.* 11 (1903): 170.

$$h = \tfrac{1}{2}\,\frac{\omega'(E)}{\omega(E)} = \frac{1}{4\kappa T}\,, \tag{1}$$ [6]

where κ denotes an absolute constant and ω is defined (slightly [7]
differently than in the article cited) by the equation

$$\omega(E).\delta E = \int_{E}^{E+\delta E} dp_1 \ldots dp_n\ .$$

The integral on the right is to be extended over all values of the state variables that completely and uniquely define the instantaneous state of the system, and to which correspond values of the energy that lie between E and $E + \delta E$.

From equation (1) it follows that

$$S = \int \frac{dE}{T} = 2\kappa \log[\omega(E)]\ .$$

Omitting the arbitrary integration constant, the expression thus represents the entropy of the system. This expression for the entropy of a system holds not only for systems that experience purely thermal changes of state, but also for systems that pass through arbitrary adiabatic and isopycnic changes of [8]
state.

The proof can be deduced from the last equation of §6, loc. cit.; I omit it because here I do not intend to present any application of the law in its general significance.

§2. *Derivation of the second law*

If a system is located in an environment of a given constant temperature T_0 and is in thermal interaction ("contact") with this environment, then, as experience shows, it too assumes the temperature T_0 and maintains this temperature T_0 for all times.

However, according to the molecular theory of heat, this law does not hold strictly, but rather in a certain approximation – even though this approximation is very good for all systems accessible to direct investigation.

If the system considered has been in that environment for an infinitely long time, the probability W that the value of the system's energy lies between the limits E and $E+1$ at an arbitrarily chosen instant (§3, *loc. cit.*) will be

[9] $$W = Ce^{-\frac{E}{2\kappa T_0}}\,\omega(E) \;,$$

where C is a constant. This value is different from zero for every E but has a maximum for a certain E and — at least for all systems accessible to direct investigation — is very small for any appreciably larger or smaller E. We call the system "heat reservoir" and assert in brief: the above expression represents the probability that the energy of the heat reservoir in question will have the value E in the environment mentioned. Using the result of the previous section, we can also write

$$W = Ce^{\frac{1}{2\kappa}\left[S - \frac{E}{T_0}\right]} \;,$$

where S denotes the entropy of the heat reservoir.

Let there be a number of heat reservoirs, all of them in the environment [10] at temperature T_0. The probability that the energy of the first reservoir will have the value E_1, the second the value $E_2 \ldots$, and the last the value E_ℓ, is, then, in an easily understood notation,

[11] (a) $$\mathfrak{W} = W_1 . W_2 \ldots W_\ell = C_1 . C_2 \ldots C_\ell e^{\frac{1}{2\kappa}\left\{\sum_1^\ell S - \frac{\sum_1^\ell E}{T_0}\right\}} \;.$$

Let these reservoirs enter into interaction with an engine that passes through a cyclic process. Assume that during this process no heat exchange takes place either between the heat reservoir and the environment or between the engine and the environment. After the process considered, let the energies and entropies of the systems be, respectively,

$$E'_1,\ E'_2 \ldots E'_\ell\ ,$$

and

$$S'_1,\ S'_2 \ldots S'_\ell\ ,$$

The probability of the total state of the heat reservoir defined by these values will be [12]

$$\text{(b)} \qquad \mathfrak{W}' = C_1 . C_2 \ldots C_\ell e^{\frac{1}{2\kappa}\left\{\sum\limits_1^\ell S' - \frac{\sum\limits_1^\ell E'}{T_0}\right\}} .$$

Neither the state of the environment nor the state of the engine has changed during the process, because the latter underwent a cyclic process.

If we now assume that less probable states never follow the more probable ones, we have

$$\mathfrak{W}' \geqq \mathfrak{W} .$$ [13]

But we also have, according to the energy principle,

$$\sum_1^\ell E = \sum_1^\ell E' .$$

If we take this into account, then it follows from equations (a) and (b) that

$$\sum S' \geqq \sum S.$$

3. On the meaning of the constant κ in the kinetic theory of atoms [14]

Let us consider a physical system whose instantaneousstate is completely determined by the values of the state variables

$$p_1, p_2 \ldots p_n \,.$$

If the system considered is in "contact" with a system of relatively infinitely large energy and of absolute temperature T_0, then its distribution of states is determined by the equation

$$dW = Ce^{-\frac{E}{2\kappa T_0}} dp_1 \ldots dp_n \,.$$

In this equation κ is a universal constant whose meaning shall now be examined.

On the basis of the kinetic theory of atoms, one arrives at an interpretation of this constant in the following way, familiar from Boltzmann's [15] works on the theory of gases.

Let the p_ν's be the orthogonal coordinates $x_1 y_1 z_1, x_2 y_2 \ldots, x_n y_n z_n$, and $\xi_1 \eta_1 \zeta_1, \xi_2 \eta_2 \ldots, \xi_n \eta_n \zeta_n$ the velocities of the individual atoms (considered to be pointlike) of the system. One can choose these state variables because [16] they satisfy the condition $\sum \partial\varphi_\nu / \partial p_\nu = 0$ (*loc. cit.*, §2). One has then:

$$E = \Phi(x_1 \ldots z_n) + \sum_1^n \frac{m_\nu}{2}(\xi_\nu^2 + \eta_\nu^2 + \zeta_\nu^2) \,,$$

where the first summand denotes the potential energy and the second the kinetic energy of the system. Let now an infinitesimally small region $dx_1 \ldots dz_n$ be given. We find the mean value of the quantity

$$\frac{m_\nu}{2}(\xi_\nu^2 + \eta_\nu^2 + \zeta_\nu^2)$$

corresponding to this region:

$$\bar{L}_\nu = \frac{m}{2}\overline{(\xi_\nu^2+\eta_\nu^2+\zeta_\nu^2)} =$$

$$= \frac{e^{-\frac{(\Phi x_1 \ldots z_n)}{4\kappa T_0}} dx_1 \ldots dz_n \int \frac{m_\nu}{2}(\xi_\nu^2+\eta_\nu^2+\zeta_\nu^2)\, e^{\frac{\sum_1^n \frac{m_\nu}{2}(\xi_\nu^2+\eta_\nu^2+\zeta_\nu^2)}{2\kappa T_0}} d\xi_1 \ldots d\zeta_n}{e^{-\frac{\Phi(x_1 \ldots z_n)}{4\kappa T_0}} dx_1 \ldots dz_n \int e^{\frac{\sum \frac{m_\nu}{2}(\xi_\nu^2+\eta_\nu^2+\zeta_\nu^2)}{2\kappa T_0}} d\xi_1 \ldots d\zeta_n} = \quad [17]$$

$$= 3\,\frac{\int_{-\infty}^{+\infty} m_\nu \xi_\nu^2 e^{\frac{m_\nu \xi_\nu^2}{4\kappa T_0}} d\xi_\nu}{\int_{-\infty}^{+\infty} e^{\frac{m_\nu \xi_\nu^2}{4\kappa T_0}} d\xi_\nu} = 3\kappa T_0 \ .$$

This quantity is thus independent of the choice of the region and the choice of the atom, and hence is in general the mean value for the atom at the absolute temperature T_0. The quantity 3κ equals the quotient of the mean kinetic energy of an atom and the absolute temperature.[1]

Further, the constant κ is closely connected with the number N of true molecules contained in one molecule as the chemists understand it (equivalent weight based on 1 g hydrogen as unit).

It is well known that for such a quantity of an ideal gas, and with gram and centimeter used as units, we have

$$pv = RT, \quad \text{where} \quad R = 8.31 \times 10^7 \ . \quad [19]$$

According to the kinetic theory of gases, however,

$$pv = \frac{2}{3} N\bar{L} \ ,$$

---

[1] Cf. L. Boltzmann, *Vorl. über Gastheorie* [Lectures on the theory of gases] *2* (1898): §42. [18]

where $\bar{L}$ denotes the mean value of the kinetic energy of motion of the center of gravity of a molecule. If one also takes into account that

$$\bar{L} = \bar{L}_{\nu}$$

one obtains

$$N \cdot 2\kappa = R \ .$$

Hence the constant 2κ equals the quotient of the constant R and the number of molecules contained in one equivalent.

[20] If, in accordance with O. E. Meyer, one sets $N = 6.4 \times 10^{23}$, one gets [21] $\kappa = 6.5 \times 10^{-17}$.

§4. *The general significance of the constant* κ

Let a given system be in contact with a system of relatively infinitely large energy and temperature T. The probability dW that the value of its energy will lie between E and $E + dE$ at an arbitrarily chosen instant is

[22]
$$dW = Ce^{-\frac{E}{2\kappa T}}\omega E dE \ .$$

For the mean value $\bar{E}$ of E one obtains

$$\bar{E} = \int_0^{\infty} CEe^{-\frac{E}{2\kappa T}}\omega E dE \ .$$

Since, further,

$$1 = \int_0^{\infty} Ce^{-\frac{E}{2\kappa T}}\omega E dE \ ,$$

we get

$$\int_0^{\infty} (\bar{E} - E)e^{-\frac{E}{2\kappa T}}\omega(E) dE = 0 \ .$$

Differentiation of this equation with respect to T yields

$$\int_0^\infty \left[2\kappa T^2 \frac{d\bar{E}}{dT} + \bar{E}E - \bar{E}^2\right] e^{-\frac{E}{2\kappa T}} \omega E dE = 0 . \qquad [23]$$

This equation states that the mean value of the bracketed expression vanishes, and hence

$$2\kappa T^2 \frac{d\bar{E}}{dT} = \bar{E}^2 - \bar{E}\bar{E} . \qquad [24]$$

In general, the instantaneous value E of the energy differs from $\bar{E}$ by a certain amount, which we call "energy fluctuation"; we put

$$E = \bar{E} + \epsilon .$$

We then obtain

$$\bar{E}^2 - \bar{E}\bar{E} = \overline{\epsilon^2} = 2\kappa T^2 \frac{d\bar{E}}{dT} . \qquad [25]$$

The quantity $\overline{\epsilon^2}$ is a measure of the thermal stability of the system; the larger the $\overline{\epsilon^2}$, the less this stability.

Thus the absolute constant κ determines the thermal stability of the system. The relationship just found is interesting because it no longer contains any quantity reminiscent of the assumptions on which the theory is based.

The magnitudes of $\overline{\epsilon^3}$, $\overline{\epsilon^4}$, etc. can be calculated by successive differentiations without any difficulty.

§5. *Application to radiation*

The last-found equation would allow an exact determination of the universal constant κ if it were possible to determine the mean value of the square of the energy fluctuation of a system; however, at the present state of our knowledge this is not the case. In fact, there is only a single kind of physical system for which we can surmise from experience that it possesses energy fluctuation: this is empty space filled with temperature radiation.

That is, if the linear dimensions of a space filled with temperature radiation are very large in comparison with the wavelength corresponding to the maximum energy of the radiation at the temperature in question, then the mean energy fluctuation will obviously be very small in comparison with the mean radiation energy of that space. In contrast, if the radiation space is of the same order of magnitude as that wavelength, then the energy fluctuation will be of the same order of magnitude as the energy of the radiation of the radiation space.

Of course, one can object that we are not permitted to assert that a radiation *space* should be viewed as a *system* of the kind we have assumed, not even if the applicability of the general molecular theory is conceded. Perhaps one would have to assume, for example, that the boundaries of the space vary with its electromagnetic states. However, these circumstances need not be considered, as we are dealing with orders of magnitude only.

If, then, in the equation obtained in the last section, we set

[26] $$\overline{\epsilon^2} = \overline{E}^2 ,$$

and according to the Stefan-Boltzmann law

[27] $$\overline{E} = cvT^4 ,$$

where v denotes the volume in cm^3 and c the constant of this law, then we must obtain for $\sqrt[3]{v}$ a value of the order of magnitude of the wavelength of the maximal radiation energy that corresponds to the temperature in question.

One obtains

$$\sqrt[3]{v} = \frac{2\sqrt[3]{\frac{\kappa}{c}}}{T} = \frac{0.42}{T} ,$$

where we have used for κ the value obtained from the kinetic theory of
[28] gases, and 7.06×10^{-15} for c.

If λ_m is the wavelength of the energy maximum of the radiation, then experiment yields

$$\lambda_m = \frac{0.293}{T} .$$ [29]

One can see that both the kind of dependence on the temperature and the order of magnitude of λ_m can be correctly determined from the general molecular theory of heat, and considering the broad generality of our assumptions, I believe that this agreement must not be ascribed to chance. [30]

Bern, 27 March 1904. (Received on 29 March 1904)

Doc. 6

Review of G. BELLUZZO, "Principles of Graphic Thermodynamics"

("Principi di termodinamica grafica," *Il Nuovo Cimento* 8 (1904):
196-222, 241-263)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 235]

This article, which is obviously meant for engineers, is divided into four sections, the first of which treats graphically the changes of state of arbitrary fluids. Thus, the familiar areal construction of the work performed (L) by the body, of the energy increase (ΔE), and of the heat absorbed (G) are given in the pv-plane in §3, while in §4 and §5 the increase of entropy for an arbitrary change of state is presented as an area with G and T (the absolute temperature), and with G and $1/T$, respectively, as coordinates. This is followed by the theory of cyclic processes and the definition of reversibility and irreversibility of the processes. A process is considered to be reversible or irreversible, respectively, depending on whether the pressure exerted on the fluid during the process does or does not equal the inner pressure of the fluid; this stipulation, which, by the way, is irrelevant for what follows, does not make sense, because then the principle of the equality of action and reaction would not be satisfied in any irreversible process. The second section of the article contains the application of the theory to ideal gases; examined are the changes of state at constant volume, constant pressure, and constant temperature, as well as the adiabatic and polytropic change of state. The last section deals with the efflux of gases through pipes; the hypothesis of Saint-Venant and Wantzel is
[1] replaced by (already known) theoretical considerations. The third and fourth sections of the article contain the theory of the saturated and the superheated water vapor, which are treated in a corresponding way, with special consideration given to the theory of the efflux of water vapor through pipes and to the theory of improving the efficiency of steam engines by superheating. For the equation of state for water vapor, $p(v + \text{const.}) =$
[2] const. T is used, following Battelli and Tumlirz.

Doc. 7

Review of A. FLIEGNER, "On Clausius's Law of Entropy" ("Über den Clausius'schen Entropiesatz," *Naturforschende Gesellschaft in Zürich. Vierteljahrsschrift* 48 (1903): 1-48)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 236]

The author examines the entropy changes of a system during a process presumed to be strictly discontinuous (discontinuous expansion of a fluid) and concludes from his calculations that the entropy decreases at the beginning of the sudden expansion. Considerations concerning irreversible chemical processes lead the author to the conclusion that the equation $dQ/T \leq dS$ holds only for exothermic but not for endothermic processes. Similarly, the equation is not supposed to hold for cooling mixtures. It is therefore understandable that the author closes with the following sentence: "Thus, the question of whether the entropy of the universe does change at all, and if it does, then in which sense, cannot yet be answered at all at present, and will probably remain undecided forever." [1]

Doc. 8

Review of W. McFadden ORR, "On Clausius' Theorem for Irreversible Cycles, and on the Increase of Entropy" (*Philosophical Magazine and Journal of Science* 8 (Series 6) (1904): 509-527)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 237]

The author shows that in the *Vorlesungen über Thermodynamik* [*Treatise on Thermodynamics*] Planck applies the concepts "reversible" and "irreversible" in [1] a sense somewhat different from that in which he defines them. Then he advances a series of objections that may be raised against various ways of [2] representing the foundations of thermodynamics; especially noteworthy among these objections is that by Bertrand, i.e., that the pressure, temperature, [3] and entropy are defined only for the case that at least sufficiently small parts of a system can be regarded as being in equilibrium; a similar objection is raised with respect to the heat supplied.

Doc. 9

Review of G. H. BRYAN, "The Law of Degradation of Energy as the Fundamental Principle of Thermodynamics" ("Das Gesetz von der Entwertung der Energie als Fundamentalprinzip der Thermodynamik," in Meyer, S., ed., *Festschrift. Ludwig Boltzmann gewidmet zum sechzigsten Geburtstage 20. Februar 1904.*
(Leipzig: J.A. Barth, 1904): 123-136)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 237]

The author starts out from the energy principle as well as the principle [1] of the decrease of free energy. The free energy (available energy) of a system is defined as the maximal mechanical work that the system can perform during changes compatible with the external conditions. This is followed by the definition of heat supplied to the system. Then the concept of thermal equilibrium, the second law, the concept of absolute temperature, and the concept of energy are developed from the stated fundamental principles in an elegant way, and, finally, the equations of thermodynamic equilibrium are derived.

Doc. 10

Review of N. N. SCHILLER, "Some Concerns Regarding the Theory of Entropy Increase Due to the Diffusion of Gases Where the Initial Pressures of the Latter Are Equal" ("Einige Bedenken betreffend die Theorie der Entropievermehrung durch Diffusion der Gase bei einander gleichen Anfangsspannungen der letzteren," in Meyer, S., ed., *Festschrift. Ludwig Boltzmann gewidmet zum sechzigsten Geburtstage 20. Februar 1904.* (Leipzig: J.A. Barth, 1904): 350-366)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 237]

First it is shown that a homogeneous gas can be reduced isothermally to an n-time smaller volume without supply of work and heat if one assumes the existence of walls that are permeable by a part of the mass of a gas but not by the rest of the mass of the gas; according to the author, this assumption [1] does not contain any contradiction. Then it is demonstrated that the expression for the entropy of a system consisting of spatially separated gases of equal temperature and pressure has the form

$$S = \left[\sum m_i R_i\right] \lg v + f(\theta) \ ; \qquad [2]$$

the entropy of the system after diffusion can be represented by the same formula. From this it is concluded that the entropy is the same before and after diffusion. The author arrives at the same result by a line of reasoning [3] that cannot be reproduced here. In this line of reasoning one operates with a surface that separates a chemically homogeneous gas into two parts such that in thermal and mechanical equilibrium the gas pressure in the two parts is different; it is (implicitly) assumed that during the passing of the gas through this surface no work is transferred to the gas by the latter.

Doc. 11

Review of J. J. WEYRAUCH, "On the Specific Heats of Superheated Water Vapor"
("Über die spezifischen Wärmen des überhitzten Wasserdampfes," *Zeitschrift*
[1] *des Vereines deutscher Ingenieure* 48 (1904): 24-28, 50-54. Reprint, 9 pp.)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 240]

Determinations made thus far of the specific heat c_p are presented and
[2] compared (I). Equations of state for water vapor suitable for practical
[3] application are presented and discussed (II) and, using those by Zeuner, c_p
and c_v for saturated steam (III) and c_p and c_v for arbitrarily superheated steam are derived thermodynamically. Then the total heat and the steam heat are determined (V). In (VI) and (VII) there follow the fundamental equations of the theory of heat for superheated steam, their application to special cases, and several numerical examples.

Doc. 12

Review of J. H. van't HOFF, "The Influence of the Changes in Specific Heat on the Work of Conversion" ("Einfluss der Änderungen der spezifischen Wärme auf die Umwandlungsarbeit," in Meyer, S., ed., *Festschrift. Ludwig Boltzmann gewidmet zum sechzigsten Geburtstage 20. Februar 1904*. (Leipzig: J.A. Barth, 1904): 233-241)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 240]

The author shows by way of thermodynamics that the work of conversion E (supplied to the surroundings) of a system A into a system B (e.g., by melting) in isothermal conversion can be represented in the form

$$E = E_0 + AT - ST \lg T.$$

(A is a constant, T the absolute temperature, $S = S_A - S_B$ the difference between the specific heats, which are assumed to be independent of T. For reasons of analogy (because during isothermal expansion of a gas $E = AT = 2T$ [1] $\lg (v_B/v_A)$), AT is regarded as determined by change of concentration.

The equation is applied to experiments of Richards, who for conversions [2] of the kind

$$Mg + ZnSO_4.aq = Zn + MgSO_4.aq$$

(where the initial $ZnSO_4$ and the $MgSO_4$ formed have the same concentration) by the electric method found that

$$\frac{dE}{dT} = -\kappa S \ ,$$

where κ is approximately the same for all conversions examined. Omitting the term AT, the author obtains from the above equation

$$\frac{dE}{dt} = - S(1 + \lg T) = - 6.7\ S\ .$$ [3]

Mean values of observations yielded: [4]

Reaction	$\left[\frac{dE}{dt}\right]/(-S)$	Reaction	$\left[\frac{dE}{dt}\right]/(-S)$
$Mg + ZnSO_4$	5	$Zn + NiSO_4$	8
$Mg + CuSO_4$	5.4	$Fe + CuSO_4$	7.5
$Mg + NiSO_4$	5.9	$Ni + CuSO_4$	7
$Mg + FeSO_4$	6.3	$Zn + CuSO_4$	7.4
$Zn + FeSO_4$	7.3	$Fe + NiSO_4$	7.1

The equation for E, applied to fusion as well as to conversion of allotropic elements and polymorphic compounds (again neglecting the term AT) further yields the proposition: The form which is stable at the higher temperature (e.g., liquid) has the higher specific heat. This conclusion is almost always confirmed by experiment. Finally, it is concluded from the

[5] equation that the Thomson-Berthelot rule must be valid at low temperatures, but that at higher temperatures the term $-ST \lg T$ may cause deviations when $S_A > S_B$.

Doc. 13

Review of A. GIAMMARCO, "A Case of Corresponding States in Thermodynamics"

("Un caso di corrispondenza in termodinamica," *Il Nuovo Cimento* 5 (5)
[1] [2] (1903): 377-391)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 246]

If one has a liquid (volume v) in a closed cylindrical tube and above it its saturated vapor (volume v'), and one plots v/v' as a function of the absolute temperature T in orthogonal coordinates, one obtains, depending on the amount of the enclosed substance, a curve that has a maximum $(v/v')_{\max}$, or a curve that is convex toward the abscissa, or one (as the limiting case) that approaches the critical temperature linearly. The author investigated ether, alcohol and chloroform in this way and finds that the above maxima $(v/v')_{\max}$ lie on a straight line. According to the law of corresponding states, two temperatures T and T' at which two different substances have the same $(v/v')_{\max}$ must be corresponding temperatures (the method for the determination of corresponding temperatures), hence $T/T_c = T'/T'_c$. Using the (absolute) critical temperatures of ether (467°), alcohol (517°), chloroform
[3] (541°) (Bureau des Longitudes, 1902), the author finds from his observations:

Corresp. abs. temperatures			$(V/V')_{\max}$			$\frac{T}{T_c}$
Ether	Alcohol	Chloroform	Ether	Alcohol	Chloroform	
387°	428.07	447.09	0.320	0.320	0.330	0.828
391	432.8	452.8	0.340	0.340	0.350	0.837
394	435.8	456	0.355	0.356	0.360	0.843
404	447	467.9	0.395	0.400	0.409	0.865
414	456.5	478	0.440	0.440	0.448	0.883
423	468.2	489.6	0.490	0.490	0.495	0.905
427	472.7	494.4	0.510	0.510	0.511	0.914
437	485.3	505.8	0.556	0.556	0.556	0.935
458	506.6	530	0.655	0.652	0.652	0.981
467	517	541	0.695	0.698	0.698	1

[4]

Examining the curve that constitutes the limiting case, the author finds that the disappearance of the meniscus during heating and its appearance during cooling occur at the same temperature (the critical temperature).

Doc. 14

ON A HEURISTIC POINT OF VIEW CONCERNING THE PRODUCTION AND TRANSFORMATION OF LIGHT [1]

by A. Einstein

[*Annalen der Physik* 17 (1905): 132-148]

There exists a profound formal difference between the theoretical conceptions physicists have formed about gases and other ponderable bodies, and Maxwell's theory of electromagnetic processes in so-called empty space. While we conceive of the state of a body as being completely determined by the positions and velocities of a very large but nevertheless finite number of atoms and electrons, we use continuous spatial functions to determine the electromagnetic state of a space, so that a finite number of quantities cannot be considered as sufficient for the complete description of the electromagnetic state of a space. According to Maxwell's theory, energy is to be considered as a continuous spatial function for all purely electromagnetic phenomena, hence also for light, while according to the current conceptions of physicists the energy of a ponderable body is to be described as a sum
[2] extending over the atoms and electrons. The energy of a ponderable body cannot be broken up into arbitrarily many, arbitrarily small parts, while according to Maxwell's theory (or, more generally, according to any wave theory) the energy of a light ray emitted from a point source of light spreads continuously over a steadily increasing volume.

The wave theory of light, which operates with continuous spatial func-
[3] tions, has proved itself splendidly in describing purely optical phenomena and will probably never be replaced by another theory. One should keep in mind, however, that optical observations apply to time averages and not to momentary values, and it is conceivable that despite the complete confirmation of the theories of diffraction, reflection, refraction, dispersion, etc., by experiment, the theory of light, which operates with continuous spatial functions, may lead to contradictions with experience when it is applied to the phenomena of production and transformation of light.

Indeed, it seems to me that the observations regarding "black-body
[4] radiation," photoluminescence, production of cathode rays by ultraviolet

light, and other groups of phenomena associated with the production or conversion of light can be understood better if one assumes that the energy of [5] light is discontinuously distributed in space. According to the assumption to be contemplated here, when a light ray is spreading from a point, the energy is not distributed continuously over ever-increasing spaces, but consists of a finite number of energy quanta that are localized in points in space, move without dividing, and can be absorbed or generated only as a whole.

In this paper I wish to communicate my train of thought and present the facts that led me to this course, in the hope that the point of view to be elaborated may prove of use to some researchers in their investigations.

§1. On a difficulty encountered in the theory of "black-body radiation"

We shall begin by taking the standpoint of Maxwell's theory and the electron theory and consider the following case. Consider a space enclosed by completely reflecting walls containing a number of gas molecules and electrons that move freely and exert conservative forces on each other when they come very close to each other, i.e., they can collide like gas molecules according to the kinetic theory of gases.[1] Suppose, further, that a number of electrons are bound to points in space which are very far from each other, by forces that are directed toward these points and are proportional to the elongations from the points. These electrons, too, shall enter into conservative interactions with the free molecules and electrons when the latter come very close to them. We call the electrons bound to the points in space "resonators"; they emit and absorb electromagnetic waves of definite periods. [7]

According to the present view about the origin of light, the radiation in the space considered, found for the case of dynamic equilibrium on the basis of Maxwell's theory, must be identical with "black-body radiation" — at [8] least if one assumes that resonators of all the relevant frequencies are present. [9]

---

[1] This assumption is equivalent to the assumption that the mean kinetic energies of gas molecules and electrons are equal to each other at thermal equilibrium. As we know, Mr. Drude used the latter assumption to derive the [6] ratio of thermal and electric conductivities of the metals theoretically.

For the time being, we disregard the radiation emitted and absorbed by the resonators and look for the condition for dynamic equilibrium corresponding to the interaction (collisions) of molecules and electrons. For such an equilibrium, the kinetic theory of gases provides the condition that the mean kinetic energy of a resonator electron must be equal to the mean kinetic energy of the progressive motion of a gas molecule. If we resolve the motion of the resonator electron into three mutually perpendicular oscillatory motions, we find for the mean value $\bar{E}$ of the energy of such a linear oscillatory motion

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

where R denotes the universal gas constant, N the number of "real [10] molecules" in one gram-equivalent, and T the absolute temperature, for because of the equality of the time averages of the resonator's kinetic and potential energies, the energy $\bar{E}$ is 2/3 times as large as the kinetic energy of a free monoatomic gas molecule. If due to some factor--in our case, due to radiation--the energy of a resonator were to have a time average larger or smaller than $\bar{E}$, the collisions with the free electrons and molecules would lead to an energy transfer to the gas or an energy absorption from the gas that is, on average, different from zero. Thus, in the case we are considering, dynamic equilibrium is possible only if the mean energy of every resonator equals $\bar{E}$.

We now apply similar reasoning to the interaction between the resonators and the radiation present in the space. Mr. Planck has derived the condition for the dynamic equilibrium in this case[1] using the assumption that

---

[11] [1] M. Planck, *Ann. d. Phys.* 1 (1900): 99.

the radiation may be considered as the most disordered process imaginable.[2] He found

$$\bar{E}_\nu = \frac{L^3}{8\pi\nu^2}\rho_\nu \; . \qquad [14]$$

$\bar{E}_\nu$ is here the mean energy of a resonator with the proper frequency ν (per oscillation component), L the velocity of light, ν the frequency, and $\rho_\nu d\nu$ the energy per unit volume of that part of the radiation whose frequency lies between ν and $\nu + d\nu$.

If, on the whole, the radiation energy of frequency ν does not continually decrease or increase, we must have

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

$$\rho_\nu = \frac{R}{N}\frac{8\pi\nu^2}{L^3}T \; . \qquad [15]$$

---

[2] This assumption can be formulated as follows. We expand the Z-component of [12]
the electrical force (Z) at an arbitrary point of the space considered in a time interval between $t = 0$ and $t = T$ (where T shall denote a time period that is very large relative to all pertinent oscillation periods) in a Fourier series

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

where $A_\nu \geq 0$ and $0 \leq a_\nu \leq 2\pi$. If one imagines that at the same point in space such an expansion is made arbitrarily often at randomly chosen initial points of time, then one will obtain different sets of values for the quantities A_ν and a_ν. For the frequency of occurrence of the various combinations of values of the quantities A_ν and a_ν, there will exist, then, (statistical) probabilities dW of the form

$$dW = f(A_1 A_2 \ldots a_1 a_2 \ldots)\, dA_1 dA_2 \ldots da_1 da_2 \ldots$$

The radiation is in the most disordered state imaginable when

$$f(A_1, A_2 \ldots a_1, a_2 \ldots) = F_1(A_1) F_2(A_2) \ldots f_1(a_1) \cdot f_2(a_2) \ldots \; ,$$

i.e., when the probability of a specific value of one of the quantities A or
a is independent of the values taken by the other quantities A and a, re- [13]
spectively. Hence, the more closely fulfilled the condition that the individual pairs of quantities A_ν and a_ν depend on the emission and absorption processes of *particular* groups of resonators, the closer to a "most disordered state imaginable" the radiation is to be viewed in our case.

This relation, obtained as the condition of dynamic equilibrium, not only [16] fails to agree with experience but it also states that in our model a definite distribution of energy between ether and matter is out of the question, since the wider the chosen range of the resonators' frequencies, the larger the radiation energy of the space, and we obtain in the limit

[17]
$$\int_0^\infty \rho_\nu d\nu = \frac{R}{N}\frac{8\pi}{L^3} T \int_0^\infty \nu^2 d\nu = \infty .$$

[18] §2. *On Planck's determination of the elementary quanta*

We now wish to show that Mr. Planck's determination of the elementary quanta is to some extent independent of his theory of "black-body radiation."

Planck's formula[1] for ρ_ν, which has been sufficient to account for all observations made so far, reads

[20]
$$\rho_\nu = \frac{\alpha\nu^3}{e^{\frac{\beta\nu}{T}} - 1} ,$$

where

$$\alpha = 6.10 \times 10^{-56}$$
$$\beta = 4.866 \times 10^{-11} .$$

For large values of T/ν, i.e., for large wavelengths and radiation densities, this formula reduces in the limit to

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

One can see that this formula agrees with that derived from the Maxwellian theory and the electron theory in §1. By equating the coefficients of the two formulas, we obtain

---

[19] [1] M. Planck, *Ann. d. Phys.* 4 (1901): 561.

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

or

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

i.e., one atom of hydrogen weighs $1/N$ gram = 1.62×10^{-24} g. This is exactly the value found by Mr. Planck, which shows satisfactory agreement with values found for this quantity by other methods. [21]

We therefore arrive at the following conclusion: the greater the energy density and the wavelength of radiation, the more useful the theoretical principles we have been using prove to be; however, these principles fail completely in the case of small wavelengths and small radiation densities.

In the following, we shall consider "black-body radiation" in connection with experience without basing it on any model for the production and propagation of radiation.

§3. *On the entropy of radiation*

The following consideration is contained in a famous study by Mr. Wien and shall be presented here only for the sake of completeness. [22]

Consider radiation that occupies a volume v. We assume that the observable properties of this radiation are completely determined when the radiation density $\rho(\nu)$ is given for all frequencies.[1] Since radiations of different frequencies are to be viewed as separable from each other without expenditure of work and without supply of heat, the entropy of radiation can be represented in the form

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

where φ is a function of the variables ρ and ν. One can reduce φ to a function of a single variable by formulating the assertion that adiabatic

---

[1] This assumption is arbitrary. Naturally, we will maintain this simplest assumption as long as the experimental results do not force us to abandon it.

compression of radiation between reflecting walls does not change its entropy. However, we shall not enter into this, but will immediately investigate how the function φ can be obtained from the black-body radiation law.

In the case of "black-body radiation," ρ is such a function of ν that the entropy is a maximum at a given energy, i.e.,

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

if

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

From this it follows that for every choice of $\delta\rho$ as function of ν

$$\int_0^\infty \left[\frac{\partial\varphi}{\partial\rho} - \lambda\right]\delta\rho\,d\nu = 0 \ ,$$

where λ is independent of ν. Thus for black-body radiation $\partial\varphi/\partial\rho$ is independent of ν.

The following equation applies when the temperature of black-body radiation of volume $v = 1$ increases by dT:

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

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

$$dS = \frac{\partial\varphi}{\partial\rho}\,dE \ .$$

Since dE equals the heat added, and the process is reversible, we also have

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

Comparison yields

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

This is the law of black-body radiation. Thus, one can determine the law of black-body radiation from the function φ, and, vice versa, the function φ can be determined by integrating the former, considering that φ vanishes for $\rho = 0$.

§4. *Limiting law for the entropy of monochromatic radiation at low radiation density*

Though the existing observations of "black-body radiation" show that the law

$$\rho = \alpha\nu^3 e^{-\beta \frac{\nu}{T}}$$ [23]

postulated by Mr. W. Wien for "black-body radiation" is not strictly valid, the law has been fully confirmed by experiment for large values of ν/T. We [24] shall base our calculations on this formula, but will keep in mind that our results are valid within certain limits only.

First of all, this formula yields

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

and next, using the relation found in the preceding section,

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

Now consider radiation of energy E whose frequency lies between ν and $\nu + d\nu$. Let the radiation occupy volume v. The entropy of this radiation is

$$S = v\varphi(\rho,\nu)d\nu = -\frac{E}{\beta\nu}\left\{\lg \frac{E}{v\alpha\nu^3 d\nu} - 1\right\} .$$ [25]

If we restrict ourselves to investigating the dependence of the entropy on the volume occupied by the radiation and denote the entropy of radiation by S_0 when the latter occupies the volume v_0, we obtain

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

This equation shows that the entropy of a monochromatic radiation of sufficiently low density varies with the volume according to the same law as [26] the entropy of an ideal gas or that of a dilute solution. The equation just found shall be interpreted in the following on the basis of the principle introduced into physics by Mr. Boltzmann, according to which the entropy of a [27] system is a function of the probability of its state.

§5. *Molecular-theoretical investigation of the dependence of the entropy of gases and dilute solutions on the volume*

In calculating the entropy by molecular-theoretical methods, the word "probability" is often used in a sense that does not coincide with the [28] definition of probability used in the probability calculus. In particular, the "cases of equal probability" are often stated hypothetically when the theoretical models applied are sufficiently definite to permit a deduction instead of a hypothetical statement. I will show in a separate paper that, when dealing with thermal processes, it is completely sufficient to use the [29] so-called "statistical probability," and I hope that this will remove a logical difficulty that still hinders the implementation of Boltzmann's principle. Here, however, I shall only give its general formulation and its application to very special cases.

If it makes sense to talk about the probability of a state of a system, and if, further, each entropy increase can be conceived as a transition to a more probable state, then the entropy S_1 of a system is a function of the probability W_1 of its instantaneous state. Therefore, if we have two systems S_1 and S_2 that do not interact with each other, we can put

$$S_1 = \varphi_1(W_1) \ ,$$
$$S_2 = \varphi_2(W_2) \ .$$

If these two systems are viewed as a single system of entropy S and probability W, we have

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

and

$$W = W_1 . W_2 \ .$$

The last relation tells us that the states of the two systems are mutually independent events.

From these equations it follows that

$$\varphi(W_1 . W_2) = \varphi_1(W_1) + \varphi_2(W_2)$$

and from this we get, finally,

$$\varphi_1(W_1) = C \ \mathrm{lg}(W_1) + \text{const.}$$
$$\varphi_2(W_2) = C \ \mathrm{lg}(W_2) + \text{const.}$$
$$\varphi(W) = C \ \mathrm{lg}(W) + \text{const.}$$

The quantity C is thus a universal constant; it follows from the kinetic theory of gases that its value is R/N, where the meaning of the constants R and N is the same as above. If S_0 denotes the entropy in some initial state of a system considered, and W the relative probability of a state having the entropy S, we obtain, in general, [30]

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

First, we deal with the following special case. Let a volume v_0 contain a number (n) of movable points (e.g., molecules), which shall be the object of our consideration. The space may also contain any number of movable points of whatever kind. No assumptions shall be made about the law governing the motion of the points in the space except that, with regard to this motion, no part of the space (and no direction) shall be distinguished from the others. The number of the (first-mentioned) movable points shall be so small that the effects of the points on each other can be disregarded.

This system, which might be, for example, an ideal gas or a diluted solution, possesses a certain entropy S_0. Let us consider a part of the volume v_0 of magnitude v and let all n movable points be transferred into the volume v without any other change in the system. It is obvious that this state has a different value of entropy (S), and we now wish to determine the entropy difference with the aid of Boltzmann's principle.

We ask: How great is the probability of the last-mentioned state relative to the original one? Or: How great is the probability that at a randomly chosen instant of time all n independently movable points in a given volume v_0 will be contained (by chance) in volume v?

Obviously, for this probability, which is a "statistical probability," one obtains the value

$$W = \left[\frac{v}{v_0}\right]^n ;$$

from this, by applying Boltzmann's principle, one obtains

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

It is noteworthy that the derivation of this equation, from which the Boyle-Gay-Lussac law and the identical law of osmotic pressure can easily be obtained by thermodynamics,[1] does not require any assumptions about the law governing the motion of the molecules.

---

[1] If E is the energy of the system, we obtain

[31] $$-d(E - TS) = pdv = TdS = R\,\frac{n}{N}\,\frac{dv}{v} ;$$

hence

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

§6. *Interpretation of the expression for the dependence of the entropy of monochromatic radiation on volume according to Boltzmann's principle*

In §4 we found the following expression for the dependence of the entropy of monochromatic radiation on volume:

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

If we write this formula in the form

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

and compare it with the general formula expressing the Boltzmann 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, the probability that at a randomly chosen instant the entire radiation energy will be contained in the portion v of the volume v_0 is

$$W = \left[\frac{v}{v_0}\right]^{\frac{N}{R}\frac{E}{\beta\nu}}$$

From this we further conclude:

Monochromatic radiation of low density (within the range of validity of Wien's radiation formula) behaves thermodynamically as if it consisted of mutually independent energy quanta of magnitude $R\beta\nu/N$.

We also wish to compare the mean value of the energy quanta of "black-body radiation" with the mean kinetic energy of the center-of-mass motion of a molecule at the same temperature. The latter is $\frac{3}{2}(R/N)T$, while the mean value of the energy quantum obtained on the basis of the Wien formula is

[32]
$$\frac{\int_0^\infty a\nu^3 e^{-\frac{\beta\nu}{T}}\,d\nu}{\int_0^\infty \frac{N}{R\beta\nu}\,a\nu^3 e^{-\frac{\beta\nu}{T}}\,d\nu} = 3\,\frac{R}{N}\,T\,.$$

If, with regard to the dependence of its entropy on volume, a monochromatic radiation (of sufficiently low density) behaves like a discontinuous medium consisting of energy quanta of magnitude $R\beta\nu/N$, then it seems reasonable to investigate whether the laws of generation and conversion of light are also so constituted as if light consisted of such energy quanta. We will now consider this question.

[33]
§7. *On Stokes' rule*

Let monochromatic light be converted by photoluminescence to light of another frequency, and let us assume in accordance with the result just obtained that both the producing and the produced light consist of energy quanta of magnitude $(R/N)\beta\nu$, where ν denotes the pertinent frequency. The conversion process is then to be interpreted as follows. Each producing energy quantum of frequency ν_1 is absorbed and—at least at a sufficiently low distribution density of the producing energy quanta—by itself gives rise to the generation of a light quantum of frequency ν_2; possibly the absorption of the producing light quantum might also be accompanied by the simultaneous generation of light quanta of frequencies ν_3, ν_4, etc., as well as of energy of some other kind (e.g., heat). It makes no difference by what kind of intermediary processes this end result is mediated. If the photoluminescent substance is not to be regarded as a permanent source of energy, then, according to the energy principle, the energy of a produced energy quantum cannot be greater than that of a producing light quantum; hence we must have

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

or

$$\nu_2 \leqq \nu_1\,.$$

This is the well-known Stokes' rule.

It should be especially emphasized that, according to our conception, at weak illumination the produced amount of light must be proportional to the intensity of the exciting light, because each exciting energy quantum will induce an elementary process of the kind indicated above, independent of the action of the other exciting energy quanta. In particular, no lower limit will exist for the intensity of the exciting light below which the light would [34] be unable to act as an exciter of light.

According to the conception of the phenomena expounded, deviations from Stokes' rule are conceivable in the following cases: [35]

1. When the number of simultaneously converting energy quanta per unit volume is so large that an energy quantum of the light produced could obtain its energy from several producing quanta;
2. When the producing (or produced) light does not have the same energy properties that obtain for "black-body radiation" within the range of validity of Wien's law as, for example, when the exciting light is produced by a body of such high temperature that Wien's law is no longer valid for the pertinent wavelength.

The latter possibility deserves special attention, for according to the conception expounded above it is not impossible that even in great dilutions the energetic behavior of a "non-Wien radiation" differs from that of a "black-body radiation" that is within the range of validity of Wien's law.

§8. *On the generation of cathode rays by illumination of solid bodies*

The usual conception, that the energy of light is continuously distributed over the space through which it travels, meets with especially great difficulties when one attempts to explain the photoelectric phenomena; these difficulties are presented in a pioneering work by Mr. Lenard.[1] [36]

According to the conception that the exciting light consists of energy quanta of energy $(R/N)\beta\nu$, the production of cathode rays by light can be conceived in the following way. The body's surface layer is penetrated by

---

[1] P. Lenard, *Ann. d. Phys.* 8 (1902): 169 and 170. [37]

energy quanta whose energy is converted at least partially to kinetic energy of electrons. The simplest possibility is that a light quantum transfers its entire energy to a single electron; we will assume that this can occur. However, we will not exclude the possibility that the electrons absorb only a part of the energy of the light quanta. An electron provided with kinetic energy in the interior of the body will have lost a part of its kinetic energy
[38] by the time it reaches the surface. In addition, it will have to be assumed that in leaving the body, each electron has to do some work P (characteristic for the body). The greatest perpendicular velocity on leaving the body will be that of electrons located directly on the surface and excited perpendicular to it. The kinetic energy of such electrons is

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

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

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

[39] where ϵ denotes the electric mass of the electron, or

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

where E denotes the charge of one gram-equivalent of a univalent ion and P' is the potential of this quantity of negative electricity with respect to the body.[1]

[41] If one sets $E = 9.6 \times 10^3$, then $\Pi . 10^{-8}$ is the potential in volts that the body acquires during irradiation in the vacuum.

To see whether the relation derived agrees with experience in order of magnitude, we put $P' = 0$, $\nu = 1.03 \times 10^{15}$ (which corresponds to the limit

---

[1] If one assumes that the release of the individual electron from a neutral
[40] molecule by light must be accompanied by the expenditure of some work, one does not have to change anything in the above relation; but then P' is to be considered as the sum of two summands.

of the solar spectrum toward the ultraviolet) and $\beta = 4.866 \times 10^{-11}$. We [42]
obtain $\Pi . 10^7 = 4.3$ volt, a result that agrees in order of magnitude with the results of Mr. Lenard.[1]

If the formula derived is correct, then Π, presented as a function of the frequency of the exciting light in Cartesian coordinates, must be a straight line whose slope is independent of the nature of the substance [44] investigated.

As far as I can see, our conception does not conflict with the properties of the photoelectric effect observed by Mr. Lenard. If each energy quantum of the exciting light transmits its energy to electrons independent of all others, then the velocity distribution of the electrons, i.e., the quality of the cathode rays produced, will be independent of the intensity of the exciting light; on the other hand, under otherwise identical circumstances, the number of electrons leaving the body will be proportional to the intensity of the exciting light.[2]

Remarks similar to those regarding the expected deviations from Stokes' rule apply to the expected limits of validity of the laws mentioned above.

In the foregoing it has been assumed that the energy of at least some of the energy quanta of the producing light is transmitted completely to one single electron each. If this obvious assumption is not made, instead of the above equation one obtains the following one:

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

For the cathode luminescence, which constitutes the inverse process of that discussed above, one obtains by a consideration analogous to that above [45]

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

For the substances investigated by Mr. Lenard, PE is always considerably [46]
larger than $R\beta\nu$ because the potential difference the cathode rays must

---

[1] P. Lenard, *Ann. d. Phys.* 8 (1902): 165 and 184, Table I, Fig. 2. [43]

[2] P. Lenard, *loc. cit.*, p. 150 and pp. 166-168.

traverse in order to produce light that is just visible amounts to several hundred volts in some cases, and to thousands of volts in others.[1] We must therefore assume that the kinetic energy of one electron is used for the production of many quanta of light energy.

[48] §9. *On the ionization of gases by ultraviolet light*

We will have to assume that in the ionization of a gas by ultraviolet light one quantum of light energy is used for the ionization of one molecule of gas. From this it follows that the work of ionization (i.e., the work theoretically required for ionization) of one molecule cannot be greater than the energy of one effective quantum of light absorbed. If J denotes the (theoretical) ionization work per gram-equivalent, we must have

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

However, according to measurements by Lenard, the largest effective wavelength [49] for air is about 1.9×10^{-5}cm, hence

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

An upper limit for the work of ionization can also be obtained from the ionization potentials in rarefied gases. According to J. Stark[2] the smallest measured ionization potential (at platinum anodes) for air is about 10 volts.[3] Thus one obtains 9.6×10^{12} as the upper limit for J, which is almost equal to the value we have just found. There is still another consequence, whose verification by experiment seems to me of great importance. If each absorbed quantum of light energy ionizes one molecule, then the following relation

---

[47] [1] P. Lenard, *Ann. d. Phys.* 12 (1903): 469.

[50] [2] J. Stark, *Die Elektrizität in Gasen*, p. 57. Leipzig, 1902

[3] In the interior of the gases the ionization potential of negative ions is five times larger, however.

must hold between the quantity of light absorbed L and the number j of gram-molecules ionized by it:

$$j = \frac{L}{R\beta\nu} \, .$$ [51]

If our conception corresponds to reality, this relation must apply to all gases that (at the relevant frequency) display no noticeable absorption that is not accompanied by ionization.

Bern, 17 March 1905. (Received on 18 March 1905)

Doc. 15

A NEW DETERMINATION OF MOLECULAR DIMENSIONS

[1] Dissertation
toward
the degree of Doctor of Philosophy
granted by
The Faculty of Philosophy
(Department of Mathematics and Natural Sciences)
of the
University of Zurich

[2] Presented
by
Albert Einstein
from Zurich

[3] Reviewed by Prof. Dr. A. Kleiner
and
Prof. Dr. H. Burkhardt

Bern

Printer K. J. Wyss
1905

[4] Dedicated to my friend
Dr. Marcel Grossmann

A NEW DETERMINATION OF MOLECULAR DIMENSIONS

The earliest determinations of real sizes of molecules were made possible by the kinetic theory of gases, whereas the physical phenomena observed in liquids have thus far not served for the determination of [5] molecular sizes. This is no doubt due to the fact that it has not yet been possible to overcome the obstacles that impede the development of a detailed molecular-kinetic theory of liquids. It will be shown in this paper that the size of molecules of substances dissolved in an undissociated dilute solution can be obtained from the internal friction of the solution and the pure solvent and from the diffusion of the dissolved substance within the solvent if the volume of the molecule of the dissolved substance is large compared with the volume of the molecule of the solvent. This is because, with respect to its mobility in the solvent and its effect on the internal friction of the latter, such a molecule will behave approximately as a solid body suspended in a solvent, and it will be permissible to apply to the motion of the solvent in the immediate vicinity of a molecule the hydrodynamic equations in which the liquid is considered to be homogeneous and hence its molecular structure is not taken into consideration. For the shape of the solid body that shall represent the dissolved molecule, we will choose the spherical shape.

§*1. On the influence on the motion of a liquid exercised by a very small sphere suspended in it*

Let us base our consideration on an incompressible homogeneous liquid with a coefficient of viscosity k, whose velocity components u, v, w are given as functions of the coordinates x, y, z and of the time. At an arbitrary point x_0, y_0, z_0, the functions u, v, w are developed as functions of $x - x_0$, $y - y_0$, $z - z_0$ according to Taylor's theorem, and around this point there is demarcated a region G that is so small that within it only the linear terms of this development must be taken into consideration. As is well known, the motion of the liquid contained in G can then be considered as a superposition of three motions, i.e.,

1. A parallel displacement of all liquid particles without a change in their relative position;

2. A rotation of the liquid without a change in the relative position of the liquid particles;

3. A dilatational motion in three mutually perpendicular directions (the principal axes of dilatation). [6]

Let us now assume that in the region G there is a spherical rigid body whose center shall lie at the point x_0, y_0, z_0 and whose dimensions shall be very small compared with those of the region G. We further assume that the motion under consideration is so slow that the kinetic energy of the sphere as well as that of the liquid can be neglected. We also assume that the velocity components of a surface element of the sphere coincide with the corresponding velocity components of the adjacent liquid particles, i.e., that the transition layer (imagined to be continuous) also displays everywhere a coefficient of viscosity that is not infinitesimally small. [7]

It is obvious that the sphere simply takes part in the partial motions 1 and 2, without modifying the motion of the neighboring particles, since the liquid moves like a rigid body in these partial motions and since we neglected the effects of inertia.

However, motion 3 does get modified by the presence of the sphere, and our next task will be to investigate the effect of the sphere on this motion of the liquid. If we refer motion 3 to a coordinate system whose axes are parallel to the principal axes of dilatation and put

$$\begin{aligned} x - x_0 &= \xi \,, \\ y - y_0 &= \eta \,, \\ z - z_0 &= \zeta \,, \end{aligned}$$

we can describe the above motion, if the sphere is not present, by the equations

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

A, B, C are constants which because of the incompressibility of the liquid satisfy the condition

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

If, now, a rigid sphere of radius P is located at point x_0, y_0, z_0, the motion of the liquid around it will change. We will, for convenience, call P "finite" and the values of ξ, η, ζ, for which the liquid motion is no longer noticeably modified by the sphere, "infinitely large."

Due to the symmetry of the motion of the liquid, it is clear that the sphere can perform neither a translation nor a rotation during the motion considered, and we obtain the boundary conditions [8]

$$u = v = w = 0 \quad \text{for} \quad \rho = P ,$$

where we have put

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

Here u, v, w denote the velocity components of the motion now considered (modified by the sphere). If we put

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

the velocities u_1, v_1, w_1 would have to vanish at infinity, since at infinity the motion represented in equations (3) should reduce to that represented by equations (1).

The functions u, v, w have to satisfy the equations of hydrodynamics including internal friction and neglecting inertia. Thus the following equations will hold[1]: [9]

---

[1] G. Kirchhoff, *Vorlesungen über Mechanik*. 26.Vorl. [Lectures on Mechanics. Lecture 26]. [10]

[11] (4)
$$\left\{\begin{array}{l} \frac{\delta p}{\delta \xi} = k\Delta u \ \frac{\delta p}{\delta \eta} = k\Delta v \ \frac{\delta p}{\delta \zeta} = \Delta w, \\ \frac{\delta u}{\delta \xi} + \frac{\delta v}{\delta \eta} + \frac{\delta w}{\delta \zeta} = 0 \ , \end{array}\right.$$

where Δ denotes the operator

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

and p the hydrostatic pressure.

Since equations (1) are solutions of equations (4) and the latter are linear, according to (3) the quantities u_1, v_1, w_1 must also satisfy equations (4). I determined u_1, v_1, w_1 and p by a method given in §4 of the Kirchhoff lectures mentioned above[1] and found

---

[12] [1] "From equations (4) it follows that $\Delta p = 0$. If we take p in keeping with this condition and determine a function V that satisfies the equation

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

then equations (4) are satisfied if one puts

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

and chooses u', v', w' such that $\Delta u' = 0$, $\Delta v' = 0$, $\Delta w' = 0$, and

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

Now, if one puts

[13]
$$\frac{p}{k} = 2c \frac{\delta^2 \frac{1}{\rho}}{\delta \xi^2}$$

and in accordance with this

[14]
$$V = c \frac{\delta^2 \rho}{\delta \xi^2} + b \frac{\delta^2 \frac{1}{\rho}}{\delta \xi^2} + \frac{a}{2} \left[\xi^2 - \frac{\eta^2}{2} - \frac{\zeta^2}{2}\right]$$

and

[15]
$$u' = -2c \frac{\delta \frac{1}{\delta}}{\delta \xi}, \qquad v' = 0, \ w' = 0,$$

then the constants a, b, c can be determined such that $u = v = w = 0$ for $\rho = P$. By superposing three such solutions, we get the solution given in equations (5) and (5a).

(5)
$$\begin{cases} p = -\frac{5}{3}kP^3\left\{A\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta\xi^2} + B\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta\eta^2} + C\,\frac{\delta^2\left[\frac{1}{\delta}\right]}{\delta\zeta^2}\right\} + \text{const.}\ , \\ u = A\xi - \frac{5}{3}P^3A\,\frac{\xi}{\rho^3} - \frac{\delta D}{\delta\xi}\ , \\ v = B\eta - \frac{5}{3}P^3B\,\frac{\eta}{\rho^3} - \frac{\delta D}{\delta\eta}\ , \\ w = C\zeta - \frac{5}{3}P^3C\,\frac{\zeta}{\rho^3} - \frac{\delta D}{\delta\zeta}\ , \end{cases}$$ [16] [17]

where

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

It can easily be proved that equations (5) are solutions of equations (4). Since

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

and

$$\Delta\left[\frac{\xi}{\rho^3}\right] = -\,\frac{\delta}{\delta\xi}\left\{\Delta\left[\frac{1}{\rho}\right]\right\} = 0\ ,$$

we get

$$k\Delta u = -\,k\,\frac{\delta}{\delta\xi}\,\{\Delta D\} = -\,k\,\frac{\delta}{\delta\xi}\left\{\frac{5}{3}\,P^3A\,\frac{\delta^2\frac{1}{\rho}}{\delta\xi^2} + \frac{5}{3}\,P^3B\,\frac{\delta^2\frac{1}{\rho}}{\delta\eta^2} + \cdots\right\}\ .$$

However, according to the first of equations (5), the last of the expressions we obtained is identical to $\frac{\delta n}{\delta\xi}$. In the same way, it can be shown that the [18] second and third of equations (4) are satisfied. Further, we get

$$\frac{\delta u}{\delta \xi} + \frac{\delta v}{\delta \eta} + \frac{\delta w}{\delta \xi} = (A + B + C) + \frac{5}{3}P^3\left\{A\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta \xi^2} + B\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta \eta^2} + C\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta \zeta^2}\right] - \Delta D\ .$$

But since according to equation (5a)

$$\Delta D = \frac{5}{3}AP^3\left\{A\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta \xi^2} + B\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta \eta^2} + C\,\frac{\delta^2\left[\frac{1}{\rho}\right]}{\delta \zeta^2}\right]\ ,$$

it follows that the last of equations (4) is satisfied as well. As far as the boundary conditions are concerned, at infinitely large ρ our equations for u, v, w reduce to equations (1). By inserting the value of D from equation (5a) into the second of equations (5), we get

[19] (6) $$u = A\xi - \frac{5}{2}\,\frac{P^3}{\rho^6}\,\xi(A\xi^2 + B\eta^2 + C\zeta^2) + \frac{5}{2}\,\frac{P^5}{\rho^7}\,\xi(A\xi^2 + B\eta^2 + C\zeta^2) - \frac{P^5}{\rho^5}\,A\xi\ .$$

We see that u vanishes for $\rho = P$. For reasons of symmetry the same holds for v and w. We have now demonstrated that equations (5) satisfy equations (4) as well as the boundary conditions of the problem.

It also can be demonstrated that equations (5) are the only solution of equations (4) compatible with the boundary conditions of our problem. We shall only indicate the proof here. Assume that in a finite space the velocity components u, v, w of a liquid satisfy equations (4). If there existed yet another solution U, V, W of equations (4) in which $U = u$, $V = v$, $W = w$ at the boundaries of the space in question, then $(U - u,\ V - v,\ W - w)$ would be a solution of equations (4) in which the velocity components vanish at the boundary of the space. Thus no mechanical work is supplied to the liquid in the space under consideration. Since we neglected the kinetic energy of the liquid, it follows that in this space the work converted to heat is also zero. This leads to the conclusion that in the entire space we must
[20] have $u = u_1$, $v = v_1$, $w = w_1$ if the space is at least partly bounded by stationary walls. By passing to the limit this result can also be extended to the case that the space under consideration is infinite as in the case considered above. This way one can show that the solution found above is the only solution of the problem.

We now place a sphere of radius R around point x_0, y_0, z_0, with R being infinitely large compared with P, and calculate the energy that is converted to heat (in unit time) in the liquid inside the sphere. This energy W is equal to the work mechanically supplied to the liquid. If X_n, Y_n, Z_n denote the components of the pressure exerted on the surface of the sphere of radius R, we have

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

where the integral is to be extended over the surface of the sphere of radius R. We have here

[21]

$$X_n = -\left[X\xi \frac{\xi}{\rho} + X\eta \, \frac{\eta}{\rho} + X\zeta \, \frac{\zeta}{\rho}\right] ,$$

$$Y_n = -\left[Y\xi \frac{\xi}{\rho} + Y\eta \, \frac{\eta}{\rho} + Y\zeta \, \frac{\zeta}{\rho}\right] ,$$

$$Z_n = -\left[Z\xi \frac{\xi}{\rho} + Z\eta \, \frac{\eta}{\rho} + Z\zeta \, \frac{\zeta}{\rho}\right] ,$$

[22]

where

$$X\xi = p - 2k \, \frac{\delta u}{\delta \xi}, \qquad Y\zeta = Z\eta = -k\left[\frac{\delta v}{\delta \zeta} + \frac{\delta w}{\delta \eta}\right] ,$$

$$Y\eta = p - 2k \, \frac{\delta v}{\delta \eta}, \qquad Z\xi = X\zeta = -k\left[\frac{\delta w}{\delta \xi} + \frac{\delta u}{\delta \zeta}\right] ,$$

$$Z\zeta = p - 2k \, \frac{\delta w}{\delta \zeta}, \qquad X\eta = Y\xi = -k\left[\frac{\delta u}{\delta \eta} + \frac{\delta v}{\delta \xi}\right] .$$

The expressions for u, v, w become simpler if we take into account that for $\rho = R$ the terms with the factor P^5/ρ^5 vanish in comparison with those with the factor P^3/ρ^3. We have to put

[23]

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

For p we obtain from the first of equations (5) by similar neglect of terms

[24] $$p = -5k\ P^3\ \frac{A\xi^2 + B\eta^2 + C\zeta^2}{\rho^5} + \text{const.}$$

First we obtain

[25] $$X\xi = -2kA + 10kP^3\ \frac{A\xi^2}{\rho^5} - 25kP^3\ \frac{\xi^2(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^7}\ ,$$

[26] $$X\eta = \qquad + 10kP^3\ \frac{A\xi\eta}{\rho^5} - 25kP^3\ \frac{\eta^2(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^7}\ ,$$

$$X\zeta = \qquad + 10kP^3\ \frac{A\xi\zeta}{\rho^5} + 25kP^3\ \frac{\zeta^2(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^7}\ .$$

and from this

[27] $$X_n = 2Ak\ \frac{\xi}{\rho} - 10AkP^3\ \frac{\xi}{\rho^4} + 25kP^3\ \frac{\xi(A\xi^2 + B\eta^2 + C\zeta^2)}{\rho^6}\ .$$

With the aid of the expressions for Y_n and Z_n derived by cyclic permutation, and neglecting all terms that contain the ratio P/ρ at a power higher than the third, we obtain

[28] $$X_n u + Y_n v + Z_n w + \frac{2k}{\rho}\ (A^2\xi^2 + B^2\eta^2 + C^2\zeta^2)$$
$$- 10k\ \frac{P^3}{\rho^4}\ (A^2\xi^2 + . + .) + 20k\ \frac{P^3}{\rho^6}\ (A\xi^2 + . + .)^2\ .$$

If we integrate over the sphere and take into account that

$$\int ds = 4R^2\pi\ ,$$

$$\int \xi^2 ds = \int \eta^2 ds = \int \zeta^2 ds = \frac{4}{3}\pi R^4\ ,$$

$$\int \xi^4 ds = \int \eta^4 ds = \int \zeta^4 ds = \frac{4}{5}\pi R^6\ ,$$

[29] $$\int \eta^2\zeta^2 ds = \int \zeta^2\xi^2 ds = \int \xi^2\eta^2 ds = \frac{4}{15}\pi R^6\ ,$$

[30] $$\int (A\xi^2 + B\eta^2 + C\zeta^2)^2 ds = \frac{4}{15}\pi R^6(A^2 + B^2 + C^2)\ ,$$

we obtain

(7) $$W = \frac{8}{3}\pi R^3 k\delta^2 - \frac{8}{3}\pi P^3 k\delta^2 = 2\delta^2 k(V - \Phi)\ ,$$ [31]

where we have put

$$\delta = A^2 + B^2 + C^2,$$ [32]

$$\frac{4}{3}\pi R^3 = V$$

and

$$\frac{4}{3}\pi P^3 = \Phi\ .$$

If the suspended sphere were not present ($\Phi = 0$), we would obtain for the energy consumed in the volume V

(7a) $$W_0 = 2\delta^2 kV\ .$$

Thus, the presence of the sphere decreases the energy consumed by $2\delta^2 k\Phi$. It is noteworthy that the effect of the suspended sphere on the quantity of [33] energy consumed is exactly the same as it would be if the presence of the sphere would not modify the motion of the liquid around it at all.

§2. *Calculation of the coefficient of viscosity of a liquid in which very many irregularly distributed small spheres are suspended*

In the previous section we have considered the case where in a region G, of the order of magnitude defined earlier, there is suspended a sphere that is very small compared with that region, and we have investigated how this sphere affects the motion of the liquid. We are now going to assume that the region G contains infinitely many randomly distributed spheres of equal radius, and that this radius is so small that the combined volume of all the spheres is very small compared with the region G. Let the number of spheres per unit volume be n, where, up to negligibly small terms, n is constant throughout the liquid.

Again, we start off from the motion of a homogeneous liquid without suspended spheres and consider again the most general motion of dilatation. If no spheres are present, an appropriate choice of the coordinate system will permit us to represent the velocity components u_0, v_0, w_0 at the arbitrary point x, y, z of the region G by the equations

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

where

$$A + B + C = 0\,.$$

[34] A sphere suspended at point x, y, z will affect this motion in the manner evident from equation (6). Since we choose the average distance between neighboring spheres to be very large compared with the radius, and since consequently the additional velocity components arising from all the suspended spheres are very small compared with u_0, v_0, w_0, we obtain for the velocity components u, v, w in the liquid, when taking into consideration the suspended spheres and neglecting terms of higher orders,

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

[35]

where the sum is to be extended over all spheres of the region G and where we have put

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

x_ν, y_ν, z_ν are the coordinates of the centers of the spheres. Furthermore, from equations (7) and (7a) we conclude that, up to infinitesimally small quantities of higher order, the presence of each sphere results in a decrease of heat production by $2\delta^2 k\Phi$ per unit time and that the energy converted to [36] heat in the region G has the value

$$W = 2\delta^2 k - 2n\delta^2 k\Phi$$

per unit volume, or

$$W = 2\delta^2 k(1 - \varphi) , \tag{7b}$$

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

Equation (7b) gives the impression that the coefficient of viscosity of [37] the inhomogeneous mixture of liquid and suspended spheres (from now on briefly called "mixture") under investigation is smaller than the coefficient of viscosity k of the liquid. However, this is not so, since A, B, C are not the values of the principal dilatations of the liquid motion represented by equations (8); we will call the principal dilatations of the mixture A^*, B^*, C^*. For reasons of symmetry, it follows that the principal dilatation directions of the mixture are parallel to the directions of the principal dilatations A, B, C, i.e., to the coordinate axes. If we write equations (8) in the form

$$\begin{aligned} u &= Ax + \sum u_\nu , \\ v &= By + \sum v_\nu , \\ z &= Cz + \sum w_\nu , \end{aligned}$$

we obtain

$$A^* = \left[\frac{\delta u}{\delta x}\right]_{x=0} = A + \sum \left[\frac{\delta u_\nu}{\delta x}\right]_{x=0} = A - \sum \left[\frac{\delta u_\nu}{\delta x_\nu}\right]_{x=0} .$$

If we exclude the immediate surroundings of the individual spheres from consideration, we can omit the second and third terms of the expressions for u, v, w and thus obtain for $x = y = z = 0$:

[38]

$$(9) \quad \left\{ \begin{aligned} u_\nu &= -\frac{5}{2}\frac{P^3}{r_\nu^2}\frac{x_\nu(Ax_\nu^2 + By_\nu^2 + Cz_\nu^2)}{r_\nu^3} , \\ v_\nu &= -\frac{5}{2}\frac{P^3}{r_\nu^2}\frac{y_\nu(Ax_\nu^2 + By_\nu^2 + Cz_\nu^2)}{r_\nu^3} \\ w_\nu &= -\frac{5}{2}\frac{P^3}{r_\nu^2}\frac{x(Ax_\nu^2 + By_\nu^2 + Cz_\nu^2)}{r_\nu^3} \end{aligned} \right.$$

where we have put

$$r_\nu = \sqrt{x_\nu^1 + y_\nu^2 + z_\nu^2} > 0 .$$

We extend the summation over the volume of a sphere K of very large radius R whose center lies in the coordinate origin. Further, if we imagine the *irregularly* distributed spheres as being *uniformly* distributed and replace the sum by an integral, we obtain

[39]

$$\begin{aligned} A^* &= A - n\int_K \frac{\delta u_\nu}{\delta x_\nu}\, dx_\nu dy_\nu dz_\nu , \\ &= A - n\int \frac{u_\nu x_\nu}{r_\nu}\, ds , \end{aligned}$$

where the last integral is to be extended over the surface of the sphere K. Taking into account (9), we find

$$\begin{aligned} A^* &= A - \frac{5}{2}\frac{P^3}{R^6}\, n\int x_0^2(Ax_0^2 + By_0^2 + Cz_0^2)\, ds , \\ &= A - n(\tfrac{4}{3} P^3\pi)A = A(1 - \varphi) . \end{aligned}$$

Analogously,

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

If we put

$$\delta^{*2} = A^{*2} + B^{*2} + C^{*2} \ ,$$ [40]

then we have, neglecting infinitesimally small terms of higher order,

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

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

$$W^* = 2\delta^2 k(1 - \varphi) \ .$$ [41]

If k^* denotes the coefficient of viscosity of the mixture, we have

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

With neglect of infinitesimal quantities of higher order, the last three equations yield

$$k^* = k(1 + \varphi) \ .$$ [42]

Thus we obtain the following result:

If very small rigid spheres are suspended in a liquid, the coefficient of internal friction increases by a fraction that is equal to the total volume [43] of the spheres suspended in unit volume, provided that this total volume is very small.

§3. *On the volume of a dissolved substance whose molecular volume is large compared with that of the solvent* [44]

Consider a dilute solution of a substance that does not dissociate in the solution. A molecule of the dissolved substance shall be large compared with a molecule of the solvent and shall be considered as a rigid sphere of radius P. In that case we can apply the result obtained in §2. If k^* denotes the coefficient of viscosity of the solution and k that of the pure solvent, we have

[45] $$\frac{k^*}{k} = 1 + \varphi ,$$

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

We wish to calculate φ for a 1% aqueous solution of sugar. According to Burkhard's observations (Landolt and Börnstein Tables), $k^*/k = 1.0245$ (at 20°C) for a 1% aqueous sugar solution, hence $\varphi = 0.0245$ for (almost exactly)
[46] 0.01 g of sugar. Thus, one gram of sugar dissolved in water has the same effect on the coefficient of viscosity as do small suspended rigid spheres of
[47] a total volume of 2.45 cm³. This consideration neglects the effect exerted on the internal friction of the solvent by the osmotic pressure resulting from the dissolved sugar.

Let us remember that 1 g of solid sugar has a volume of 0.61 cm³. This same volume is also found for the specific volume s of sugar in solution if one considers the sugar solution as a *mixture* of water and sugar in dissolved form. I.e., the density of a 1% aqueous sugar solution (referred to water of
[48] the same temperature) at 17.5° is 1.00388. Hence we have (neglecting the difference between the density of water at 4° and at 17.5°)

$$\frac{1}{1.00388} = 0.99 + 0.01\, s ,$$

and thus

$$s = 0.61 .$$

Thus, while the sugar solution behaves as a mixture of water and solid
[49] sugar with respect to its density, the effect on internal friction is four times larger than that which would result from the suspension of the same amount of sugar. It seems to me that from the point of view of the molecular theory, this result can hardly be interpreted otherwise than by assuming that the sugar molecule in the solution impedes the mobility of the water in its immediate vicinity, so that an amount of water whose volume is about three

times larger than the volume of the sugar molecule is attached to the sugar molecule. [50]

Hence we may say that a dissolved molecule of sugar (i.e., the molecule together with the water held by it) behaves in the hydrodynamic sense as a sphere with a volume of $2.45 \cdot 342/N$ cm^3, where 342 is the molecular weight of sugar and N is the number of real molecules in one gram-molecule. [51]

§4. *On the diffusion of an undissociated substance in a liquid solution*

Let us consider a solution of the kind considered in §3. If a force K acts upon the molecule, which we consider as a sphere of radius P, the molecule will move with a velocity ω, which is determined by P and the coefficient of viscosity k of the solvent, since we have the equation[1]

$$\omega = \frac{K}{6\pi kP} \,. \tag{1}$$

We use this relation to calculate the coefficient of diffusion of an undissociated solution. If p denotes the osmotic pressure of the dissolved substance, which is to be regarded as the only motion-producing force in the dilute solution under consideration, then the force acting in the direction of the X-axis on the dissolved substance per unit volume of the solution equals $-\delta p/\delta x$. If there are ρ grams in a unit volume, and m is the molecular weight of the dissolved substance and N the number of real molecules in one gram-molecule, then $(\rho/m)\cdot N$ is the number of (real) molecules in the unit volume, and the force exerted on the molecule by virtue of the concentration gradient is [53]

$$K = -\frac{m}{\rho N}\frac{\delta p}{\delta x} \,. \tag{2}$$

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

---

[1] G. Kirchhoff, *Vorlesungen über Mechanik*. 26.Vorl. [Lectures on Mechanics, Lecture 26], equation (22). [52]

[54] (3)
$$p = \frac{R}{m} \rho T \,,$$

where T is the absolute temperature and $R = 8.31 \cdot 10^7$. From equations (1), (2) and (3) we obtain for the migration velocity of the dissolved substance

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

Finally, the amount of substance passing per unit time through a unit cross section in the direction of the X-axis is

(4)
$$\omega \rho = - \frac{RT}{6\pi k} \cdot \frac{1}{NP} \frac{\delta \rho}{\delta x} \,.$$

Hence we obtain for the coefficient of diffusion D

[55]
$$D = \frac{RT}{6\pi k} \cdot \frac{1}{NP} \,.$$

Thus, from the coefficient of diffusion and the coefficient of viscosity of the solvent we can calculate the product of the number N of real molecules in one gram-molecule and the hydrodynamically effective molecular radius P.

In this derivation the osmotic pressure has been treated as a force acting on the individual molecules, which obviously does not agree with the point of view of the kinetic molecular theory, since according to the latter the osmotic pressure in the case under consideration has to be conceived as an apparent force only. However, this difficulty disappears when one considers that the (apparent) osmotic forces which correspond to the concentration differences in the solution may be kept in (dynamic) equilibrium with numerically equal forces acting on the individual molecules in the opposite direction, which can easily be realized on the basis of thermodynamics.

The osmotic force acting on the unit mass $-\frac{1}{\rho}\frac{\delta p}{\delta x}$ can be counterbalanced by the force $-P_x$ (exerted on the individual dissolved molecules) if

$$- \frac{1}{\rho} \frac{\delta p}{\delta x} - P_x = 0 \,.$$

Thus, if one imagines that the dissolved substance per unit mass is acted upon by two force systems P_x and $-P_x$ that balance each other, then $-P_x$ counterbalances the osmotic pressure and only the force P_x, which is numerically equal to the osmotic pressure, remains as the cause of motion. The difficulty mentioned above has thus been eliminated.[1]

§5. *Determination of the molecular dimensions with the help of the relations obtained*

We found in §3

$$\frac{k^*}{k} = 1 + \varphi = 1 + n \cdot \frac{4}{3} \pi P^3 ,$$ [57]

where n is the number of dissolved molecules per unit volume and P is the hydrodynamically effective radius of the molecule. If we take into account that

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

where ρ denotes the mass of the dissolved substance per unit volume and m its molecular weight, we obtain

$$NP^3 = \frac{3}{4\pi} \frac{m}{\rho} \left[\frac{k^*}{k} - 1\right] .$$ [58]

On the other hand, we found in §4 that

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

These two equations enable us to calculate separately the quantities P and N, of which N must turn out to be independent of the nature of the solvent, the dissolved substance, and the temperature, if our theory corresponds to the facts.

---

[1] A detailed presentation of this line of reasoning can be found in *Ann. d. Phys.* 17 (1905): 549. [56]

We will carry out the calculation for an aqueous solution of sugar. From the data on the internal friction of the sugar solution cited earlier, it follows that for 20°C

[59] $$NP^3 = 200 \ .$$

According to experiments of Graham (calculated by Stefan), the [60] coefficient of diffusion of sugar in water is 0.384 at 9.5°C, if the day is [61] chosen as the unit of time. The viscosity of water at 9.5° is 0.0135. We will insert these data in our formula for the coefficient of diffusion, even though they have been obtained using 10% solutions, and a strict validity of our formula cannot be expected at such high concentrations. We obtain

$$NP = 2.08 \cdot 10^{16} \ .$$

Neglecting the differences between the values of P at 9.5° and 20°, the values found for NP^3 and NP yield

[62]
$$P = 9.9 \times 10^{-8} \text{ cm} \ ,$$
$$N = 2.1 \times 10^{23} \ .$$

The value found for N shows satisfactory agreement in its order of [63] magnitude with values found for this quantity by other methods.

Bern, 30 April 1905.

Doc. 16

ON THE MOVEMENT OF SMALL PARTICLES SUSPENDED IN STATIONARY LIQUIDS REQUIRED BY THE MOLECULAR-KINETIC THEORY OF HEAT

by A. Einstein

[*Annalen der Physik* 17 (1905): 549-560]

It will be shown in this paper that, according to the molecular-kinetic theory of heat, bodies of microscopically visible size suspended in liquids must, as a result of thermal molecular motions, perform motions of such magnitude that these motions can easily be detected by a microscope. It is possible that the motions to be discussed here are identical with the so-called "Brownian molecular motion"; however, the data available to me on the latter are so imprecise that I could not form a definite opinion on this matter. [1]

If it is really possible to observe the motion to be discussed here, along with the laws it is expected to obey, then classical thermodynamics can no longer be viewed as strictly valid even for microscopically distinguishable spaces, and an exact determination of the real size of atoms becomes possible. Conversely, if the prediction of this motion were to be proved wrong, this fact would provide a weighty argument against the molecular-kinetic conception of heat. [2]

§*1*. *On the osmotic pressure attributable to suspended particles*

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

$$pV^* = RTz.$$ [3]

But if instead of the dissolved substance, the partial volume V^* of the liquid contains small suspended bodies that likewise 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—that a force be exerted on the wall; because according to the customary conception, 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 masses and properties of the suspended substance, the [4] 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 in the calculation of the free energy; but we can disregard them since the changes in the position of the wall and the suspended bodies considered here shall proceed without changes in the size and condition of the contact surfaces.

But from the standpoint of the molecular-kinetic theory of heat we are led to a different conception. According to this theory, a dissolved molecule differs from a suspended body in size *alone*, and it is difficult to see why suspended bodies should not produce the same osmotic pressure as an equal number of dissolved molecules. We will have to assume that the suspended bodies perform an irregular, even though 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 forces upon the wall exactly as dissolved molecules do. Thus, if n suspended bodies are present in the volume V^*, i.e., $n/V = \nu$ in the unit volume, and if the separation between neighboring bodies is sufficiently large, there will correspond to them an osmotic pressure p of magnitude

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

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

§2. *Osmotic pressure from the standpoint of the molecular-kinetic theory of heat*[1]

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

$$\frac{\partial p_\nu}{\partial t} = \varphi_\nu(p_1 \ldots p_\ell) \quad (\nu = 1,2,\ldots \ell),$$

where $\sum \frac{\partial \varphi_\nu}{\partial p_\nu} = 0$, then the entropy of the system is given by the expression [6]

$$S = \frac{\bar{E}}{T} + 2\kappa \lg \int e^{-\frac{E}{2\kappa T}} dp_1 \ldots dp_\ell.$$ [7]

Here T denotes the absolute temperature, $\bar{E}$ the energy of the physical system, and E the energy as a function of the p_ν's. The integral is to be extended over all combinations of values of p_ν consistent with the conditions of the problem. κ is connected with the constant N mentioned above [8] by the relation $2\kappa N = R$. We therefore get for the free energy F

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

Let us now imagine a liquid enclosed in the volume V; let the partial volume V^* of V contain n dissociated molecules or suspended bodies, which are retained in the volume V^* by a semipermeable wall; this will affect the integration limits of the integral B entering the expressions for S and F. Let the total volume of the dissolved molecules or suspended bodies

---

[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): 120). Knowledge of the papers cited and of this section of [5] the present paper is not essential for the understanding of the present paper's results.

be small compared with V^*. In accordance with the theory mentioned, this system shall be completely described by the state variables $p_1 \ldots p_\ell$.

Even if the molecular picture were established down to the smallest detail, the calculation of the integal B would be so difficult as to make an exact calculation of F all but inconceivable. However, here we only have to know how F depends on the size of the volume V^* in which all the dissolved molecules or suspended bodies (hereafter briefly called "particles") are contained.

Let us denote by x_1, y_1, z_1 the rectangular coordinates of the center of gravity of the first particle, by x_2, y_2, z_2 those of the second, etc., and by x_n, y_n, z_n those of the last particle, and assign to the centers of gravity of the particles the infinitesimally small parallelepiped-shaped regions $dx_1 dy_1 dz_1$, $dx_2 dy_2 dz_2 \ldots dx_n dy_n dz_n$, all of which shall lie in V^*. We now seek the value of 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

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

where J is independent of $dx_1 dy_1$, etc., as well as of V^*, i.e., of the position of the 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 we will show immediately. For if a second system of infinitesimally small regions were assigned to the centers of gravity of the particles and denoted by $dx_1' dy_1' dz_1'$, $dx_2' dy_2' dz_2' \ldots dx_n' dy_n' dz_n'$, 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

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

where

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

Hence,

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

But from the molecular theory of heat, presented in the papers cited[1], it can easily be deduced that dB/B and dB'/B are 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 $(dx'_1 \dots dz'_n)$, respectively. If the motions of the individual particles are (in sufficient approximation) independent of each other, and the liquid is homogeneous and no forces act upon the particles, then the probabilities corresponding to the two systems of regions must be the same if the size of the regions is the same, so that we have

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

But it follows from this equation and the one preceding it that

$$J = J'.$$

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

$$B = \int J dx_1 \dots dz_n = JV^{*n},$$

and from that

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

and

$$p = -\frac{\partial F}{\partial V^*} = \frac{RT}{V^*}\frac{n}{N} = \frac{RT}{N}\nu.$$ [10]

This consideration demonstrates that the existence of osmotic pressure is a consequence of the molecular-kinetic theory of heat, and that, according to this theory, at great dilutions numerically equal quantities of dissolved molecules and suspended particles behave completely identically with regard to osmotic pressure.

---

[1] A. Einstein, *Ann. d. Phys.* 11 (1903): 170. [9]

[11] §3. *Theory of diffusion of small suspended spheres*

Suppose that suspended particles are randomly distributed in a liquid. We wish to 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 will assume that the force is everywhere in the direction of the X-axis.

If the number of suspended particles per unit volume is ν, then in the case of thermodynamic 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. Thus

[12]
$$\delta F = \delta E - T\delta S = 0.$$

Let us assume that the liquid has a cross section 1 perpendicular to the X-axis, and that it is bounded by the planes $x = 0$ and $x = \ell$. We then have

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

and

[13]
$$\delta S = \int_0^\ell R \frac{\nu}{N} \frac{\partial \delta x}{\partial x} dx = - \frac{R}{N} \int_0^\ell \frac{\partial \nu}{\partial x} \delta x dx.$$

Hence, the equilibrium condition sought is

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

or

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

The last equation states that the force K is balanced by the forces of osmotic pressure.

We use equation (1) to determine the coefficient of diffusion of the suspended substance. The state of dynamic equilibrium that we have just considered can be conceived as a superposition of two processes proceeding in opposite directions, namely, [14]

1. a motion of the suspended substance under the influence of the force K which is exerted on each suspended particle,

2. a process of diffusion, which is to be conceived as the result of the random motions of the particles due to thermal molecular motion.

If the suspended particles are of spherical shape (where P is the radius of the sphere) and the coefficient of friction of the liquid is k, then the force K imparts to the individual particle the velocity[1]

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

and

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

particles pass through the unit cross section per unit time.

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

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

or

$$-D\,\frac{\partial\nu}{\partial x}$$

particles will pass through the unit cross section per unit time due to diffusion. Since there should be dynamic equilibrium, we must have

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

From the two conditions (1) and (2) found for dynamic equilibrium we can calculate the coefficient of diffusion. We obtain

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

---

[1] Cf., e.g., G. Kirchhoff, "*Vorlesungen über Mechanik*" [Lectures on Mechanics], Lecture 26, §4. [15]

Thus, apart from universal constants and the absolute temperature, the coefficient of diffusion of the suspended substance depends only on the coefficient of friction of the liquid and the size of the suspended particles.

§4. *On the random motion of particles suspended in a liquid and their relation to diffusion*

We shall now turn to a closer examination of the random motions which, caused by thermal molecular motion, give rise to the diffusion investigated in the last section.

Obviously, we must assume that each individual particle performs a motion that is independent of the motions of all the other particles; similarly, the motions of one and the same particle in different time intervals will have to be conceived as mutually independent processes so long as we think of these time intervals as chosen not to be too small.

We now introduce into the consideration a time interval τ, which shall be very small compared with observable time intervals but still so large that the motions performed by a particle during two consecutive time intervals τ [16] may be considered as mutually independent events.

Suppose, now, that a total of n 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 frequency law will hold for Δ: the number dn of particles experiencing a displacement lying between Δ and $\Delta + d\Delta$ in the time interval τ will be expressed by an equation of the form

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

where

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

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

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

Now we investigate how the coefficient of diffusion depends on φ, restricting ourselves again to the case that the number ν of particles per unit volume depends only on x and t.

Let $\nu = f(x,t)$ be the number of particles per unit volume; we then calculate the distribution of the 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 $x+dx$. We obtain

$$f(x,t+\tau)dx = dx.\int_{\Delta=-\infty}^{\Delta=+\infty} f(x+\Delta)\varphi(\Delta)d\Delta.$$ [17]

But since τ is very small, we can put

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

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

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

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

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

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

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

putting

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

and only considering the first and third term of the right-hand side:

[18] (1) $$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2} .$$

This is the familiar differential equation for diffusion, and D can be recognized as the diffusion coefficient.

Another important consideration can be linked to this development. We assumed that all the individual particles are 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 at time $t = 0$ with the position of the center of gravity of the particle in question, 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 lying between x and $x + dx$. Thus, the function f varies according to equation (1) in this case as well. Further, it is obvious that for $x \gtrless 0$ and $t = 0$ we must have

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

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

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

The frequency distribution of the changes of position occurring in the arbitrary time t is thus the same as the distribution of random errors,
[19] which was to be expected. What is of importance, however, is how the constant in the exponent is related to the coefficient of diffusion. With the help of this equation we 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

square root of the arithmetic mean of the squares of displacements in the direction of the X-axis; we get

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

The mean displacement is thus proportional to the square root of 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 $\lambda_x\sqrt{3}$.

§5. *Formula for the mean displacement of suspended particles. A new method of determining the true size of atoms*

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

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

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

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

Eliminating D, we get

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

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

We now wish to calculate the magnitude of λ_x for one second if N is
taken to be $6\cdot 10^{23}$ in accordance with the results of the kinetic theory of [20]
gases; water at 17°C ($k = 1.35\cdot 10^{-2}$) shall be chosen as the liquid, and the [21]
diameter of the particles shall be 0.001 mm. We obtain [22]

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

Thus, the mean displacement in 1 min. would be about 6 microns.

Conversely, the relation found can be used for the determination of N. We obtain

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

Let us hope that a researcher will soon succeed in solving the problem [23] posed here, which is of such importance in the theory of heat!

Bern, May 1905. (Received on 11 May 1905)

Doc. 17

Review of K. F. SLOTTE, "On the Heat of Fusion" ("Über die Schmelzwärme," *Finska Vetenskaps-Societeten. Öfversigt af Förhandlingar* 47, no. 7 (1904): 1-8)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 623]

The author bases himself on a relation between the heat of fusion ℓ, the (absolute) melting temperature T_1, and the specific heat at constant temperature c_p, which he had derived earlier by elementary considerations based on molecular theory, and which is expressed approximately by the formula $\ell = 0.382 c_p T_1$. This formula proves to be valid with rough approximation for [1] elements as well as for compounds. A few substances for which the formula does not hold even approximately (sulfur, phosphorus) are also presented. Incidentally, it should be noted that up to the numerical value of the constant, the relation presented, extended to the solid aggregation state, is a consequence of the law of corresponding states. Finally, a molecular-theoretical consideration, which shall not be presented here, leads the author to the view that one obtains the best agreement of the theory with experience when one ascribes linear harmonic oscillations to the atoms of simple solids. [2]

Doc. 18

Review of K. F. SLOTTE, "Conclusions Drawn from a Thermodynamic Equation" ("Folgerungen aus einer thermodynamischen Gleichung," *Finska Vetenskaps-Societeten. Öfversigt af Förhandlingar* 47, no. 8 (1904): 1-3)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 629]

Several conclusions regarding the behavior of bodies near the absolute zero of the temperature are drawn from the familiar equation

$$\left[\frac{dQ}{dv}\right]_T = T\left[\frac{dp}{dT}\right]_v$$ [1]

under the totally unexplained assumption that $(dQ/dv)_T$ has finite values at infinitesimally small values of T. [2]

Doc. 19

[1] Review of M. E. MATHIAS, "The Constant a of Rectilinear Diameters and the Laws of Corresponding States" ("La constante a des diamètres rectilignes et les lois des états correspondents (2e mémoire)," *Journal de Physique théorique et appliquée* 4 (Series 4) (1905): 77-91)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 634]

[2] If y denotes a function of the density of a liquid and its saturated vapor that depends linearly on the temperature, which has already been examined by the author in previous articles (*J. de Phys.* (3) 8 (1899): 407,
[3] and *ibid.* (3) 2 (1893): 5), then the relation $y = \Delta(1 + a[1 + m])$ holds, where Δ denotes the critical density, m the temperature, with the critical temperature taken as the unit, and a a constant. If the law of corresponding
[4] states were strictly fulfilled, a would have to be a universal constant. Based on experimental data on 37 substances it is shown that this is not the case. While for the majority of the examined substances a deviates only little from unity, this quantity has considerably smaller values for gases difficult to liquefy, the value for hydrogen being 0.236. The author finds now that even though the quantity $b = a/\sqrt{\theta}$ (θ = absolute critical temperature) is not a universal constant either, it has nevertheless almost the same value for substances of similar chemical constitution; he proposes that substances be divided into "series" (substances with almost equal b) and "groups" (substances with approximately equal a).

Doc. 20

Review of M. PLANCK, "On Clausius' Theorem for Irreversible Cycles and on the Increase of Entropy" (*Philosophical Magazine and Journal of Science* 9 (Series 6) (1905): 167-168)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 635]

In response to some objections raised by Mr. Orr (*Beibl.* 29, p. 237) [1]
against the treatment of the fundamental laws of thermodynamics presented by
the author, the author explains that he used the expressions "reversible" and
"irreversible" in the same sense as Clausius. He denies that he applied the [2]
above concepts in a way that differed from that in which he defined them. The
author admits that one cannot talk about the temperature and density of any
small parts of a tumultuously moving gas, and about their entropy, unless one
wants to make use of the kinetic theory of gases. Finally, he finds that the [3]
line of proof proposed by Mr. Orr coincides in principle with that presented
by Lord Kelvin and that it contains circular reasoning. [4]

Doc. 21

Review of E. BUCKINGHAM, "On Certain Difficulties Which are Encountered in the Study of Thermodynamics" (*Philosophical Magazine and Journal of Science* 9 (Series 6) (1905): 208-214)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 635]

The author starts from an article by Mr. Orr (*Beibl.* 29, p. 237) and [1]
expresses his agreement with the results of the critical considerations contained in that article. Further, he gives expression to his conviction that it is impossible to derive the Clausius inequality $\int dQ/T < 0$ from the second law—as formulated by Lord Kelvin—without further assumptions. The article also contains some critical remarks on Mr. Orr's paper.

Doc. 22

Review of P. LANGEVIN, "On a Fundamental Formula of the Kinetic Theory"
("Sur une formule fondamentale de la théorie cinétique," *Academie des Sciences (Paris). Comptes rendus* 140 (1905): 35-38)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 640]

The author reports that, assuming arbitrary laws of action between
[1] molecules as well as external forces acting upon the molecules, he has solved
[2] exactly the problem of diffusion of two gases by the Maxwell-Kirchhoff method,
[3] requiring only a graphic integration. For the case that the molecules are
elastic spheres which are only infinitesimally deformable, and that external
forces do not act on the molecules, the author obtains for the diffusion of
one gas (molecular mass m_1) in the other gas (molecular mass m)

[4]
$$D = \frac{3}{16\sigma^2 N\sqrt{\frac{\pi h m m_1}{m + m_1}}} .$$

Here D denotes the diffusion constant, σ the sum of the radii of two unlike molecules, N the number of molecules "m" per unit volume, h three-quarters of the reciprocal of the mean value of the energy of the translational motion of one molecule. Boltzmann found by the Clausius
[5] approximation method

[6]
$$D = \frac{2}{3\pi\sigma^2 N\sqrt{\pi h(m + m_1)}} .$$

The two formulas differ especially strongly when m and m_1 are very different. It is further reported that at constant pressure the diffusion coefficient varies as $T^{3/2 + 2/n}$ when two unlike molecules repel each other with a force that is inversely proportional to the n + 1st power of the distance between the centers of the molecules. The author has also applied the theory to changes in position of electric charges in gases. He found that the assumption of polarizing forces exerted by the electrical particle on the neutral molecules does not suffice for the explanation of their small mobility, but that, in dry air and at normal temperature, one must ascribe to the negative ions a diameter about twice as large, and to positive ions one about three times as large as that ascribed to the neutral molecules. For flames, the author finds that it is to be concluded from the empirical results that the mass of the negative electricity carriers is about a thousand times smaller than that of the positive ones, and that the mass of the latter equals that of the hydrogen atom; hence, the former corresponds to cathode rays, the latter to Goldstein rays. [7]

Doc. 23

ON THE ELECTRODYNAMICS OF MOVING BODIES

by A. Einstein

[*Annalen der Physik* 17 (1905): 891-921]

It is well known that Maxwell's electrodynamics—as usually understood at present—when applied to moving bodies, leads to asymmetries that do not seem to attach to the phenomena. Let us recall, for example, the electro-
[1] dynamic interaction between a magnet and a conductor. The observable phenomenon depends here only on the relative motion of conductor and magnet, while according to the customary conception the two cases, in which, respectively, either the one or the other of the two bodies is the one in motion, are to be strictly differentiated from each other. For if the magnet is in motion and the conductor is at rest, there arises in the surroundings of the magnet an electric field endowed with a certain energy value that produces a current in the places where parts of the conductor are located. But if the magnet is at rest and the conductor is in motion, no electric field arises in the surroundings of the magnet, while in the conductor an electromotive force will arise, to which in itself there does not correspond any energy, but which, provided that the relative motion in the two cases considered is the same, gives rise to electrical currents that have the same magnitude and the same course as those produced by the electric forces in the first-mentioned case.

Examples of a similar kind, and the failure of attempts to detect a
[2] motion of the earth relative to the "light medium", lead to the conjecture that not only in mechanics, but in electrodynamics as well, the phenomena do not have any properties corresponding to the concept of absolute rest, but that in all coordinate systems in which the mechanical equations are valid, also the same electrodynamic and optical laws are valid, as has already been
[3] shown for quantities of the first order. We shall raise this conjecture
[4] (whose content will be called "the principle of relativity" hereafter) to the status of a postulate and shall introduce, in addition, the postulate, only seemingly incompatible with the former one, that in empty space light is always propagated with a definite velocity V which is independent of the
[5] state of motion of the emitting body. These two postulates suffice for arriving at a simple and consistent electrodynamics of moving bodies on the

basis of Maxwell's theory for bodies at rest. The introduction of a "light ether" will prove superfluous, inasmuch as in accordance with the concept to [6] be developed here, no "space at absolute rest" endowed with special properties will be introduced, nor will a velocity vector be assigned to a point of empty space at which electromagnetic processes are taking place. [7]

Like every other electrodynamics, the theory to be developed is based on the kinematics of the rigid body, since assertions of each and any theory concern the relations between rigid bodies (coordinate systems), clocks, and electromagnetic processes. Insufficient regard for this circumstance is at the root of the difficulties with which the electrodynamics of moving bodies must presently grapple.

I. Kinematic Part

§*1. Definition of simultaneity*

Consider a coordinate system in which the Newtonian mechanical equations are valid. To distinguish it verbally from the coordinate systems that will [8] be introduced later on, and to visualize it more precisely, we will designate this system as the "system at rest."

If a material point is at rest relative to this coordinate system, its position relative to the latter can be determined by means of rigid measuring rods using the methods of Euclidean geometry and can be expressed in Cartesian coordinates.

If we want to describe the *motion* of a material point, we give the values of its coordinates as a function of time. However, we should keep in mind that for such a mathematical description to have physical meaning, we first have to clarify what is to be understood here by "time." We have to bear in mind that all our propositions involving time are always propositions about *simultaneous events*. If, for example, I say that "the train arrives here at 7 o'clock," that means, more or less, "the pointing of the small hand of my clock to 7 and the arrival of the train are simultaneous events."[1]

---

[1] We shall not discuss here the imprecision that is inherent in the concept of simultaneity of two events taking place at (approximately) the same location and that also must be surmounted by an abstraction.

It might seem that all difficulties involved in the definition of "time" could be overcome by my substituting "position of the small hand of my clock" for "time." Such a definition is indeed sufficient if time has to be defined exclusively for the place at which the clock is located; but the definition becomes insuffícent as soon as series of events occurring at different locations have to be linked temporally, or—what amounts to the same—events occurring at places remote from the clock have to be evaluated temporally. [9]

To be sure, we could content ourselves with evaluating the time of the events by stationing an observer with the clock at the coordinate origin, and having him assign the corresponding clock-hand position to each light signal that attests to an event to be evaluated and reaches him through empty space. But as we know from experience, such an assignment has the drawback that it is not independent of the position of the observer equipped with the clock. We arrive at a far more practical arrangement by the following consideration. [10]

If there is a clock at point A of space, then an observer located at A can evaluate the time of the events in the immediate vicinity of A by finding the clock-hand positions that are simultaneous with these events. If there is also a clock at point B—we should add, "a clock of exactly the same constitution as that at A"—then the time of the events in the immediate vicinity of B can likewise be evaluated by an observer located at B. But it is not possible to compare the time of an event at A with one at B without a further stipulation; thus far we have only defined an "A-time" and a "B-time" but not a "time" common to A and B. The latter can now be determined by establishing *by definition* that the "time" needed for the light to travel from A to B is equal to the "time" it needs to travel from B to A. For, suppose a ray of light leaves from A toward B at "A-time" t_A, is reflected from B toward A at "B-time" t_B, and arrives back at A at "A-time" t'_A. The two clocks are synchronous by definition if

$$t_B - t_A = t'_A - t_B.$$

We assume that it is possible for this definition of synchronism to be free of contradictions, and to be so for arbitrarily many points, and that the following relations are therefore generally valid:

1. If the clock in *B* is synchronous with the clock in *A*, then the clock in *A* is synchronous with the clock in *B*.

2. If the clock in *A* is synchronous with the clock in *B* as well as with the clock in *C*, then the clocks in *B* and *C* are also synchronous relative to each other.

With the help of some physical (thought) experiments, we have thus laid down what is to be understood by synchronous clocks at rest that are situated at different places, and have obviously obtained thereby a definition of "synchronous" and of "time." The "time" of an event is the reading obtained simultaneously with the event from a clock at rest that is located at the place of the event and that for all time determinations is in synchrony with a specified clock at rest.

Based on experience, we also postulate that the quantity

$$\frac{2\overline{AB}}{t'_A - t_A} = V$$

is a universal constant (the velocity of light in empty space).

It is essential that we have defined time by means of clocks at rest in a system at rest; because it belongs to the system at rest, we designate the time just defined as "the time of the system at rest."

§2. *On the relativity of lengths and times*

The considerations that follow are based on the principle of relativity and the principle of the constancy of the velocity of light, two principles that we define as follows: [11]

1. The laws governing the changes of the state of any physical system do not depend on which one of two coordinate systems in uniform translational motion relative to each other these changes of the state are referred to. [12]

2. Each ray of light moves in the coordinate system "at rest" with the definite velocity *V* independent of whether this ray of light is emitted by a body at rest or a body in motion. Here, [13]

$$\text{velocity} = \frac{\text{light path}}{\text{time interval}} ,$$

where "time interval" should be understood in the sense of the definition in §1.

Let there be given a rigid rod at rest; its length, measured by a measuring rod that is also at rest, shall be ℓ. We now imagine that the axis of the rod is placed along the X-axis of the coordinate system at rest, and that the rod is then set in uniform parallel translational motion (velocity v) along the X-axis in the direction of increasing x. We now seek to determine the length of the *moving* rod, which we imagine to be obtained by the following two operations:

(a) The observer co-moves with the above-mentioned measuring rod and the rod to be measured, and measures the length of the rod directly, by applying the measuring rod exactly as if the rod to be measured, the observer, and the measuring rod were at rest.

(b) Using clocks at rest that are set up in the system at rest and are synchronous in the sense of §1, the observer determines the points of the system at rest at which the beginning and the end of the rod to be measured are found at some given time t. The distance between these two points, measured by the rod used before, which in the present case is at rest, is also a length, which can be designated as the "length of the rod."

According to the principle of relativity, the length to be found in operation (a), which we shall call "the length of the rod in the moving system," must equal the length ℓ of the rod at rest.

We will determine the length to be found in operation (b), which we shall call "the length of the (moving) rod in the system at rest," on the basis of our two principles, and will find it to be different from ℓ.

The commonly used kinematics tacitly assumes that the lengths determined by the two methods mentioned are exactly identical, or, in other words, that in the time epoch t a moving rigid body is totally replaceable, as far as geometry is concerned, by the *same* body when it is *at rest* in a particular position.

Further, we imagine that the two ends (A and B) of the rod are equipped with clocks that are synchronous with the clocks of the system at rest, i.e., whose readings always correspond to the "time of the system at rest" at the locations they happen to occupy; hence, these clocks are "synchronous in the system at rest."

We further imagine that each clock has an observer co-moving with it, and that these observers apply to the two clocks the criterion for synchronism formulated in §1. Suppose a ray of light starts out from A at time[1] t_A, is reflected from B at time t_B, and arrives back at A at time t'_A. Taking into account the principle of the constancy of the velocity of light, we find that

$$t_B - t_A = \frac{r_{AB}}{V - v}$$

and

$$t'_A - t_B = \frac{r_{AB}}{V + v},$$

where r_{AB} denotes the length of the moving rod, measured in the system at rest. The observers co-moving with the moving rod would thus find that the two clocks do not run synchronously while the observers in the system at rest would declare them synchronous.

Thus we see that we must not ascribe *absolute* meaning to the concept of simultaneity; instead, two events that are simultaneous when observed from some particular coordinate system can no longer be considered simultaneous when observed from a system that is moving relative to that system.

§3. *Theory of transformation of coordinates and time from a system at rest to a system in uniform translational motion relative to it*

Let there be given two coordinate systems in the space "at rest," i.e., two systems of three mutually perpendicular rigid material lines issuing from one point. Let the X-axes of the two systems coincide and their Y- and Z-axes be parallel. Each system shall be supplied with a rigid measuring rod and a number of clocks, and the two measuring rods and all the clocks of the two systems should be exactly alike.

---

[1] "Time" here means both "time of the system at rest" and "the position of the hands of the moving clock located at the place in question."

The origin of one of the two systems (k) shall now be imparted a (constant) velocity v in the direction of increasing x of the other system (K), which is at rest, and this velocity shall also be imparted to the
[14] coordinate axes, the corresponding measuring rod, and the clocks. To each time t of the system at rest K there corresponds then a definite position of the axes of the moving system, and for reasons of symmetry we may rightfully assume that the motion of k can be such that at time t ("t" always denotes a time of the system at rest) the axes of the moving system are parallel to the axes of the system at rest.

We now imagine the space to be measured both from the system at rest K by means of the measuring rod at rest and from the moving system k by means of the measuring rod moving along with it, and that the coordinates x, y, z and ξ, η, ζ are obtained in this way. Further, by means of the clocks at rest in the system at rest and using light signals in the manner described in §1, the time t of the system at rest is determined for all its points where there is a clock; likewise, the time τ of the moving system is determined for all the points of the moving system having clocks that are at rest relative to this system, applying the method of light signals described in §1 between the points containing these clocks.

To every system of values x, y, z, t that determines completely the place and time of an event in the system at rest, there corresponds a system of values ξ, η, ζ, τ that fixes this event relative to the system k, and the problem to be solved is to find the system of equations connecting these quantities.

First of all, it is clear that these equations must be *linear* because of the properties of homogeneity that we attribute to space and time.

If we put $x' = x - vt$, then it is clear that a point at rest in the system k has a definite, time-independent system of values x', y, z belonging to it. We first determine τ as a function of x', y, z, and t. To this end, we must express in equations that τ is in fact the aggregate of the readings of the clocks at rest in the system k, which have been synchronized according to the rule given in §1.

Suppose that at time τ_0 a light ray is sent from the origin of the system k along the X-axis to x' and is reflected from there at time τ_1 toward the origin, where it arrives at time τ_2; we then must have

$$\frac{1}{2}(\tau_0 + \tau_2) = \tau_1,$$

or, if we write out the arguments of the function τ and apply the principle of the constancy of the velocity of light in the system at rest,

$$\frac{1}{2}\left[\tau(0,0,0,t) + \tau\left[0,0,0,\left\{t + \frac{x'}{V - v} + \frac{x'}{V + v}\right\}\right]\right] =$$

$$= \tau\left[x',0,0,t + \frac{x'}{V - v}\right].$$

From this we get, if x' is chosen infinitesimally small,

$$\frac{1}{2}\left[\frac{1}{V - v} + \frac{1}{V + v}\right]\frac{\partial\tau}{\partial t} = \frac{\partial\tau}{\partial x'} + \frac{1}{V - v}\frac{\partial\tau}{\partial t},$$

or

$$\frac{\partial\tau}{\partial x'} + \frac{v}{V^2 - v^2}\frac{\partial\tau}{\partial t} = 0.$$

It should be noted that, instead of the coordinate origin, we could have chosen any other point as the starting point of the light ray, and the equation just derived therefore holds for all values of x', y, z.

Analogous reasoning—applied to the H and Z axes—yields, if we consider that light always propagates along these axes with the velocity $\sqrt{V^2 - v^2}$ when observed from the system at rest,

$$\frac{\partial\tau}{\partial y} = 0$$

$$\frac{\partial\tau}{\partial z} = 0.$$

These equations yield, since τ is a *linear* function,

$$\tau = a\left[t - \frac{v}{V^2 - v^2}x'\right],$$

where a is a function $\varphi(v)$ as yet unknown, and where we assume for brevity that at the origin of k we have $t = 0$ when $\tau = 0$.

Using this result, we can easily determine the quantities ξ, η, ζ by expressing in equations that (as demanded by the principle of the constancy of the velocity of light in conjunction with the principle of relativity) light propagates with velocity V also when measured in the moving system. For a light ray emitted at time $\tau = 0$ in the direction of increasing ξ, we will have

$$\xi = V\tau,$$

or

$$\xi = aV\left[t - \frac{v}{V^2 - v^2}\, x'\right].$$

But as measured in the system at rest, the light ray propagates with velocity $V - v$ relative to the origin of k, so that

$$\frac{x'}{V - v} = t.$$

Substituting this value of t in the equation for ξ, we obtain

$$\xi = a\,\frac{V^2}{V^2 - v^2}\, x'.$$

Analogously, by considering light rays moving along the two other axes, we get

$$\eta = V\tau = aV\left[t - \frac{v}{V^2 - v^2}\, x'\right],$$

where

$$\frac{y}{\sqrt{V^2 - v^2}} = t; \quad x' = 0;$$

hence

$$\eta = \frac{V}{\sqrt{V^2 - v^2}}\, y$$

and

$$\zeta = \frac{V}{\sqrt{V^2 - v^2}}\, z.$$

If we substitute for x' its value, we obtain

$$\tau = \varphi(v)\beta\left[t - \frac{v}{V^2}x\right],$$

$$\xi = \varphi(v)\beta(x - vt),$$

$$\eta = \varphi(v)y,$$

$$\zeta = \varphi(v)z,$$

where

$$\beta = \frac{1}{\sqrt{1 - \left[\frac{v}{V}\right]^2}}$$

and φ is a function of v that is as yet unknown. If no assumptions are made regarding the initial position of the moving system and the zero point of τ, then an additive constant must be attached to the right-hand sides of these equations.

Now we have to prove that every light ray measured in the moving system propagates with the velocity V, if it does so, as we have assumed, in the system at rest; for we have not yet provided the proof that the principle of the constancy of the velocity of light is compatible with the relativity principle.

Suppose that at time $t = \tau = 0$ a spherical wave is emitted from the coordinate origin, which is at that time common to the two systems, and that this wave propagates in the system K with the velocity V. Hence, if (x,y,z) is a point just reached by this wave, we will have

$$x^2 + y^2 + z^2 = V^2t^2.$$

We transform these equations using our transformation equations, and, after a simple calculation, obtain

$$\xi^2 + \eta^2 + \zeta^2 = V^2\tau^2.$$

Thus, the wave under consideration is a spherical wave of propagation velocity V also when it is observed in the moving system. This proves that our two fundamental principles are compatible. [15]

The transformation equations we have derived also contain an unknown function φ of v, which we now wish to determine.

To this end we introduce a third coordinate system K', which relative to the system k is in parallel-translational motion parallel to the axis Ξ such that its origin moves along the Ξ-axis with velocity $-v$. Let all three coordinate origins coincide at time $t = 0$, and let the time t' of the system K' be zero at $t = x = y = z = 0$. We denote the coordinates measured in the system K' by x', y', z' and, by twofold application of our transformation equations, we get

$$\begin{aligned}
t' &= \varphi(-v)\beta(-v)\left\{\tau + \frac{v}{V^2}\xi\right\} &&= \varphi(v)\varphi(-v)t,\\
x' &= \varphi(-v)\beta(-v)\{\xi + v\tau\} &&= \varphi(v)\varphi(-v)x,\\
y' &= \varphi(-v)\eta &&= \varphi(v)\varphi(-v)y,\\
z' &= \varphi(-v)\zeta &&= \varphi(v)\varphi(-v)z.
\end{aligned}$$

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 relative to each other, and it is clear that the transformation from K to K' must be the identity transformation. Hence,

$$\varphi(v)\varphi(-v) = 1.$$

Let us now explore the meaning of $\varphi(v)$. We shall focus on that portion of the H-axis of the system k that lies between $\xi = 0$, $\eta = 0$, $\zeta = 0$, and $\xi = 0$, $\eta = \ell$, $\zeta = 0$. This portion of the H-axis is a rod that moves perpendicular to its axis with a velocity v relative to the system K and whose ends possess in K the coordinates

$$x_1 = vt, \quad y_1 = \frac{\ell}{\varphi(v)}, \quad z_1 = 0$$

and

$$x_2 = vt, \quad y_2 = 0, \quad z_2 = 0.$$

The length of the rod, measured in K, is thus $\ell/\varphi(v)$; this establishes the meaning of the function φ. For reasons of symmetry it is obvious that the length of a rod measured in the system at rest and moving perpendicular to its own axis can depend only on its velocity and not on the direction and sense of its motion. Thus, the length of the moving rod measured in the system at rest does not change when v is replaced by $-v$. From this we arrive at

$$\frac{\ell}{\varphi(v)} = \frac{\ell}{\varphi(-v)},$$

or

$$\varphi(v) = \varphi(-v).$$

It follows from this relation and the one found before that $\varphi(v)$ must equal 1, so that the transformation equations obtained become [16]

$$\tau = \beta\left[t - \frac{v}{V^2}x\right],$$

$$\xi = \beta(x - vt),$$

$$\eta = y,$$

$$\zeta = z,$$

where

$$\beta = \sqrt{1 - \left[\frac{v}{V}\right]^2}.$$

§4. *The physical meaning of the equations obtained concerning moving rigid bodies and moving clocks*

We consider a rigid sphere[1] of radius R that is at rest relative to the moving system k and whose center lies at the origin of k. The equation of the surface of this sphere, which moves with velocity v relative to the system k, is [17]

---

[1] I.e., a body possessing the shape of a sphere when investigated at rest.

$$\xi^2 + \eta^2 + \zeta^2 = R^2.$$

Expressed in x, y, z, the equation of this surface at time $t = 0$ is

$$\frac{x^2}{\left[\sqrt{1 - \left(\frac{v}{V}\right)^2}\right]^2} + y^2 + z^2 = R^2.$$

A rigid body that has a spherical shape when measured in the state of rest thus in the state of motion—observed from a system at rest—has the shape of an ellipsoid of revolution with axes

$$R\sqrt{1 - \left(\frac{v}{V}\right)^2}, R, R.$$

Thus, while the Y and Z dimensions of the sphere (and hence also of every rigid body, whatever its shape) do not appear to be altered by motion, the X dimension appears to be contracted in the ratio $1 : \sqrt{1 - (v/V)^2}$, i.e., the greater the value of v, the greater the contraction. At $v = V$, all moving objects—observed from the system "at rest"—shrink into plane structures. For superluminal velocities our considerations become meaningless; we shall see in the considerations that follow that in our theory the velocity of light physically plays the part of infinitely great velocities.

It is clear that the same results apply for bodies at rest in a system "at rest" that are observed from a uniformly moving system.

We further imagine that one of the clocks that is able to indicate time t when at rest relative to the system at rest and time τ when at rest relative to the system in motion, is placed in the origin of k and set such that it indicates the time τ. What is the rate of this clock when observed from the system at rest?

The quantities x, t, and τ, which refer to the position of this clock, are obviously related by the equations

$$\tau = \frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} \left(t - \frac{v}{V^2}\,x\right)$$

and

$$x = vt.$$

We thus have

$$\tau = t\sqrt{1 - \left(\frac{v}{V}\right)^2} = t - \left(1 - \sqrt{1 - \left(\frac{v}{V}\right)^2}\right)t,$$

which shows that the clock (observed in the system at rest) is retarded each second by $(1 - \sqrt{1 - v/V^2})$ sec or, with quantities of the fourth and higher orders neglected, by $\frac{1}{2}(v/V)^2$ sec.

This yields the following peculiar consequence: If at the points A and B of K there are located clocks at rest which, observed in a system at rest, are synchronized, and if the clock in A is transported to B along the connecting line with velocity v, then upon arrival of this clock at B the two clocks will no longer be synchronized; instead, the clock that has been transported from A to B will lag $\frac{1}{2}tv^2/V^2$ sec (up to quantities of the fourth and higher orders) behind the clock that has been in B from the outset, if t is the time needed by the clock to travel from A to B.

We see at once that this result holds even when the clock moves from A to B along any arbitrary polygonal line, and even when the points A and B coincide. [18]

If we assume that the result proved for a polygonal line holds also for a continuously curved line, then we arrive at the following proposition: If there are two synchronous clocks in A, and one of them is moved along a closed curve with constant velocity until it has returned to A, which takes, say, t sec, then this clock will lag on its arrival at A $\frac{1}{2}t(v/V)^2$ sec behind the clock that has not been moved. From this we conclude that a [19] balance-wheel clock that is located at the Earth's equator must be very slightly slower than an absolutely identical clock, subjected to otherwise identical conditions, that is located at one of the Earth's poles.

§5. *The addition theorem of velocities*

In the system k moving with velocity v along the X-axis of the system K let there be a point moving according to the equation

$$\xi = w_\xi \tau,$$

$$\eta = w_\eta \tau,$$

$$\zeta = 0,$$

where w_ξ and w_η denote constants.

We seek the motion of the point relative to the system K. Introducing the quantities x, y, z, t into the equations of motion of the point by means of the transformation equations derived in §3, we obtain

$$x = \frac{w_\xi + v}{1 + \frac{v w_\xi}{V^2}} t,$$

$$y = \frac{\sqrt{1 - \left(\frac{v}{V}\right)^2}}{1 + \frac{v w_\xi}{V^2}} w_\eta t,$$

$$z = 0.$$

Thus, according to our theory, the law of the parallelogram of velocities holds only in first approximation. We put

$$U^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2,$$

$$w^2 = w_\xi^2 + w_\eta^2$$

and

[20] $$\alpha = \operatorname{arctg} \frac{w_y}{w_x};$$

α should then be considered as the angle between the velocities v and w. After a simple calculation, we obtain

$$U = \frac{\sqrt{(v^2 + w^2 + 2vw \cos \alpha) - \left[\frac{vw \sin \alpha}{V}\right]^2}}{1 + \frac{vw \cos \alpha}{V^2}}.$$

It is noteworthy that v and w enter the expression for the resultant velocity in a symmetric fashion. If w too has the direction of the X-axis (Ξ-axis), we obtain

$$U = \frac{v + w}{1 + \frac{vw}{V^2}} .$$

It follows from this equation that the composition of two velocities that are smaller than V always results in a velocity that is smaller than V. For if we put $v = V - \kappa$, and $w = V - \lambda$, where κ and λ are positive and smaller than V, we get

$$U = V \frac{2V - \kappa - \lambda}{2V - \kappa - \lambda + \frac{\kappa\lambda}{V}} < V.$$

It follows further that the velocity of light V cannot be changed by compounding it with a "subluminal velocity." For this case we get

$$U = \frac{V + w}{1 + \frac{w}{v}} = V.$$

For the case that v and w have the same direction, the formula for U could also have been obtained by compounding two transformations according to §3. If in addition to the systems K and k, which figure in §3, we also introduce a third coordinate system k', which moves parallel to k and whose origin moves with velocity w along the axis Ξ, we obtain relations between the quantities x, y, z, t and the corresponding quantities of k' that differ from those found in §3 only insofar as "v" is being replaced by the quantity

$$\frac{v + w}{1 + \frac{vw}{V^2}} ;$$

from this we see that such parallel transformations form a group—as they
[21] indeed must.

We have now derived the required propositions of the kinematics that corresponds to our two principles, and will now proceed to show their application in electrodynamics.

II. Electrodynamic Part

§6. Transformation of the Maxwell-Hertz equations for empty space. On the nature of the electromotive forces that arise upon motion in a magnetic field

Let the Maxwell-Hertz equations for empty space be valid for the system
[22] at rest K, so that we have

$$\frac{1}{V}\frac{\partial X}{\partial t}=\frac{\partial N}{\partial y}-\frac{\partial M}{\partial z}\,, \qquad \frac{1}{V}\frac{\partial L}{\partial t}=\frac{\partial Y}{\partial z}-\frac{\partial Z}{\partial y}\,,$$

$$\frac{1}{V}\frac{\partial Y}{\partial t}=\frac{\partial L}{\partial z}-\frac{\partial N}{\partial x}\,, \qquad \frac{1}{V}\frac{\partial M}{\partial t}=\frac{\partial Z}{\partial x}-\frac{\partial X}{\partial z}\,,$$

$$\frac{1}{V}\frac{\partial Z}{\partial t}=\frac{\partial M}{\partial x}-\frac{\partial L}{\partial y}\,, \qquad \frac{1}{V}\frac{\partial N}{\partial t}=\frac{\partial X}{\partial y}-\frac{\partial Y}{\partial x}\,,$$

where (X,Y,Z) denotes the vector of the electric force and (L,M,N) that of the magnetic force.

If we apply the transformations derived in §3 to these equations in that we refer the electromagnetic processes to the coordinate system introduced there, which moves with velocity v, we obtain the following equations:

$$\frac{1}{V}\frac{\partial X}{\partial \tau}=\frac{\partial \beta\left[N-\frac{v}{V}Y\right]}{\partial \eta}-\frac{\partial \beta\left[M+\frac{v}{V}Z\right]}{\partial \zeta}\,,$$

$$\frac{1}{V}\frac{\partial \beta\left[Y-\frac{v}{V}N\right]}{\partial \tau}=\frac{\partial L}{\partial \zeta}-\frac{\partial \beta\left[N-\frac{v}{V}Y\right]}{\partial \xi}\,,$$

$$\frac{1}{V}\frac{\partial\beta\left[Z+\frac{v}{V}M\right]}{\partial\tau}=\frac{\partial\beta\left[M+\frac{v}{V}Z\right]}{\partial\xi}-\frac{\partial L}{\partial\eta},$$

$$\frac{1}{V}\frac{\partial L}{\partial\tau}=\frac{\partial\beta\left[Y-\frac{v}{V}N\right]}{\partial\zeta}-\frac{\partial\beta\left[Z+\frac{v}{V}M\right]}{\partial\eta},$$

$$\frac{1}{V}\frac{\partial\beta\left[M+\frac{v}{V}Z\right]}{\partial\tau}=\frac{\partial\beta\left[Z+\frac{v}{V}M\right]}{\partial\xi}-\frac{\partial X}{\partial\zeta},$$

$$\frac{1}{V}\frac{\partial\beta\left[N-\frac{v}{V}Y\right]}{\partial\tau}=\frac{\partial X}{\partial\eta}-\frac{\partial\beta\left[Y-\frac{v}{V}N\right]}{\partial\xi},$$

where

$$\beta=\frac{1}{\sqrt{1-\left[\frac{v}{V}\right]^2}}\,.$$

The relativity principle demands that the Maxwell-Hertz equations for empty space also be valid in the system k if they are valid in the system K, i.e., 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 in this system by their ponderomotive effects on the electric and magnetic masses, respectively, satisfy the equations

$$\frac{1}{V}\frac{\partial X'}{\partial\tau}=\frac{\partial N'}{\partial\eta}-\frac{\partial M'}{\partial\zeta}, \qquad \frac{1}{V}\frac{\partial L'}{\partial\tau}=\frac{\partial Y'}{\partial\zeta}-\frac{\partial Z'}{\partial\eta},$$

$$\frac{1}{V}\frac{\partial Y'}{\partial\tau}=\frac{\partial L'}{\partial\zeta}-\frac{\partial N'}{\partial\xi}, \qquad \frac{1}{V}\frac{\partial M'}{\partial\tau}=\frac{\partial Z'}{\partial\xi}-\frac{\partial X'}{\partial\zeta},$$

$$\frac{1}{V}\frac{\partial Z'}{\partial\tau}=\frac{\partial M'}{\partial\xi}-\frac{\partial L'}{\partial\eta}, \qquad \frac{1}{V}\frac{\partial N'}{\partial\tau}=\frac{\partial X'}{\partial\eta}-\frac{\partial Y'}{\partial\xi}.$$

Obviously, the two systems of equations found for the system k must express exactly the same thing, since both are equivalent to the Maxwell-Hertz equations for the system K. Further, since the equations of the two systems coincide apart from the symbols representing the vectors, it follows that the functions occurring in the systems of equations at corresponding places must coincide up to a possibly v-dependent factor $\psi(v)$, which is common to all functions of one system of equations and is independent of ξ, η, ζ, and τ. The following relations will therefore be valid:

$$X' = \psi(v)X, \qquad L' = \psi(v)L,$$

$$Y' = \psi(v)\beta\left[Y - \frac{v}{V}N\right], \qquad M' = \psi(v)\beta\left[M + \frac{v}{V}Z\right],$$

$$Z' = \psi(v)\beta\left[Z + \frac{v}{V}M\right], \qquad N' = \psi(v)\beta\left[N - \frac{v}{V}Y\right].$$

If we now invert this system of equations, first, by solving the equations just obtained and, second, by applying the equations to the inverse transformation (from k to K) which is characterized by the velocity $-v$, we obtain, if we take into account that the two systems of equations so obtained must be identical,

$$\psi(v).\psi(-v) = 1.$$

Further, it follows for reasons of symmetry[1] that

$$\psi(v) = \psi(-v);$$

thus

$$\psi(v) = 1,$$

and our equations take the form

$$X' = X, \qquad L' = L,$$

$$Y' = \beta\left[Y - \frac{v}{V}N\right], \qquad M' = \beta\left[M + \frac{v}{V}Z\right],$$

$$Z' = \beta\left[Z + \frac{v}{V}M\right], \qquad N' = \beta\left[N - \frac{v}{V}Y\right].$$

---

[1] If, e.g., $X = Y = Z = L = M = 0$ and $N \neq 0$, then it is clear for reasons of symmetry that if v changes its sign without changing its numerical value, then Y' too must change its sign without changing its numerical value.

By way of interpreting these equations, we shall add the following: Imagine a pointlike quantity of electricity whose magnitude, measured in the system at rest K, is "one," i.e., which, when at rest in the system at rest, exerts a force of 1 dyne on an equal quantity of electricity at a distance of 1 cm. [23] According to the principle of relativity this electric mass is also of magnitude "one" if measured in a moving system. If this quantity of electricity is at rest relative to the system at rest, the vector (X,Y,Z) equals the force exerted on it by definition. If this quantity of electricity is at rest relative to the moving system (at least at the instant considered), the force exerted on it, and measured in the moving system, will equal the vector (X',Y',Z'). Hence, the first three of the above equations can be expressed in words in the following two ways:

1. If a pointlike unit electric pole is in motion in an electromagnetic field, there will act on it, in addition to the electric force, an "electromotive force" which, if we neglect terms multiplied by the second and higher [24] powers of v/V, equals the vector product of the velocity of motion of the unit pole and the magnetic force, divided by the velocity of light. (Old mode of expression.)

2. If a pointlike unit electric pole is in motion in an electromagnetic field, the force acting on it equals the electric force present at the location of the unit pole, which is obtained by transforming the field to a coordinate system that is at rest relative to the unit electric pole. (New mode of expression.)

Analogous propositions apply for "magnetomotive forces." We can see [25] that in the theory developed, the electromotive force merely plays the role of an auxiliary concept, whose introduction is due to the circumstance that the electric and magnetic forces do not have an existence independent of the state of motion of the coordinate system.

It is further clear that the asymmetry mentioned in the Introduction [26] when considering the currents produced by the relative motion of a magnet and a conductor, disappears. Questions as to the "seat" of the electrodynamic electromotive forces (unipolar machines) also become pointless. [27]

§7. *Theory of Doppler's principle and of aberration*

Imagine in the system K, very far from the coordinate origin, a source of electrodynamic waves, which in a part of space containing the coordinate origin is represented with sufficient accuracy by the equations

$$\begin{aligned} X &= X_0 \sin\phi, & L &= L_0 \sin\phi, \\ Y &= Y_0 \sin\phi, & M &= M_0 \sin\phi, & \qquad \phi &= \omega\left[t - \frac{ax + by + cz}{V}\right] \\ Z &= Z_0 \sin\phi, & N &= N_0 \sin\phi. \end{aligned}$$

Here (X_0, Y_0, Z_0) and (L_0, M_0, N_0) are the vectors determining the amplitude of the wave train, and a, b, c are the direction cosines of the wave normals.

We ask, what characterizes these waves when investigated by an observer who is at rest in the moving system k? — Applying the transformation equations for electric and magnetic forces found in §6 and those for coordinates and time found in §3, we obtain directly

$$\begin{aligned} X' &= X_0 \sin\phi', & L' &= L_0 \sin\phi', \\ Y' &= \beta\left[Y_0 - \frac{v}{V} N_0\right]\sin\phi', & M' &= \beta\left[M_0 + \frac{v}{V} Z_0\right]\sin\phi', \\ Z' &= \beta\left[Z_0 + \frac{v}{V} M_0\right]\sin\phi', & N' &= \beta\left[N_0 - \frac{v}{V} Y_0\right]\sin\phi', \end{aligned}$$

$$\phi' = \omega'\left[t - \frac{a'\xi + b'\eta + c'\zeta}{V}\right],$$

where we have put

$$\omega' = \omega\beta\left[1 - a\frac{v}{V}\right],$$

$$a' = \frac{a - \frac{v}{V}}{1 - a\frac{v}{V}},$$

$$b' = \frac{b}{\beta\left[1 - a\frac{v}{V}\right]},$$

$$c' = \frac{c}{\beta\left[1 - a\frac{v}{V}\right]} .$$

From the equation for ω' it follows that if an observer moves with velocity v relative to an infinitely distant source of light of frequency ν, such that the connecting line "light source - observer" forms an angle φ with the observer's velocity, where this velocity is referred to a coordinate system that is at rest relative to the light source, then ν', the frequency of the light perceived by the observer, is given by the equation

$$\nu' = \nu \frac{1 - \cos\varphi \frac{v}{V}}{\sqrt{1 - \left[\frac{v}{V}\right]^2}} .$$

This is Doppler's principle for arbitrary velocities. For $\varphi = 0$ the equation takes the simple form

$$\nu' = \nu \sqrt{\frac{1 - \frac{v}{V}}{1 + \frac{v}{V}}} .$$

We see that, contrary to the usual conception, when $v = -\infty$, then $\nu = \infty$. [28]

If φ' denotes the angle between the wave normal (the direction of the ray) in the moving system and the connecting line "light source - observer," the equation for a' takes the form [29]

$$\cos\varphi' = \frac{\cos\varphi - \frac{v}{V}}{1 - \frac{v}{V}\cos\varphi} .$$

This equation expresses the law of aberration in its most general form. If $\varphi = \pi/2$, the equation takes the simple form

$$\cos\varphi' = -\frac{v}{V} .$$

It remains now to find the amplitude of the waves the way it appears in the moving system. If A and A' denote the electric or magnetic force in the system at rest and in motion, respectively, we get

[30]

$$A'^2 = A^2 \frac{\left[1 - \frac{v}{V} \cos \varphi\right]^2}{1 - \left[\frac{v}{V}\right]^2},$$

which for $\varphi = 0$ simplifies to

$$A'^2 = A^2 \frac{1 - \frac{v}{V}}{1 + \frac{v}{V}}.$$

It follows from the equations derived above that to an observer approaching a light source with velocity V, this source would appear to have infinite intensity.

§8. *Transformation of the energy of light rays.* *Theory of the radiation pressure exerted on perfect mirrors.*

Since $A^2/8\pi$ equals the energy of light per unit volume, according to the principle of relativity we have to consider $A'^2/8\pi$ as the light energy in the moving system. Hence A'^2/A^2 would be the ratio of the energy of a [31] given light complex "measured in motion" and the same energy "measured at rest," if the volume of a light complex were the same whether measured in K or k. However, this is not the case. If a,b,c are the direction cosines of the wave normal of the light in the system at rest, then the surface elements of the spherical surface

$$(x - Vat)^2 + (y - Vbt)^2 + (z - Vct)^2 = R^2,$$

which moves with the velocity of light, are not traversed by any energy; we may therefore say that this surface permanently encloses the same light complex. We ask for the quantity of energy enclosed by this surface as observed in the system k, i.e., the energy of the light complex relative to the system k.

Observed in the moving system, the spherical surface is an ellipsoidal surface whose equation at time $\tau = 0$ is

$$\left[\beta\xi - a\beta \frac{v}{V} \xi\right]^2 + \left[\eta - b\beta \frac{v}{V} \xi\right]^2 + \left[\zeta - c\beta \frac{v}{V}\right]^2 = R^2.$$

If S denotes the volume of the sphere and S' that of the ellipsoid, then a simple calculation shows that

$$\frac{S'}{S} = \frac{\sqrt{1 - \left(\frac{v}{V}\right)^2}}{1 - \frac{v}{V} \cos \varphi}.$$

If the energy of the light enclosed by the surface under consideration is denoted by E when measured in the system at rest and by E' when measured in the moving system, we obtain

$$\frac{E'}{E} = \frac{\frac{A'^2}{8\pi} S'}{\frac{A^2}{8\pi} S} = \frac{1 - \frac{v}{V} \cos \varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}},$$

which for $\varphi = 0$ reduces to the simpler formula

$$\frac{E'}{E} = \sqrt{\frac{1 - \frac{v}{V}}{1 + \frac{v}{V}}}.$$

It is noteworthy that the energy and the frequency of a light complex vary with the observer's state of motion according to the same law.

Let the coordinate plane $\xi = 0$ be a completely reflecting surface at which the plane waves considered in the last section are getting reflected. We ask for the light pressure exerted on the reflecting surface and the direction, frequency, and intensity of the light after reflection. [32]

Let the incident light be defined by the quantities A, cos φ, and ν (referred to the system K). Observed from k, the corresponding quantities are

$$A' = A\,\frac{1 - \frac{v}{V}\cos\varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}}\,,$$

$$\cos\varphi' = \frac{\cos\varphi - \frac{v}{V}}{1 - \frac{v}{V}\cos\varphi}\,,$$

$$\nu' = \nu\,\frac{1 - \frac{v}{V}\cos\varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}}\,.$$

Referring the process to the system k, we get for the reflected light

$$A'' = A',$$

$$\cos\varphi'' = -\cos\varphi',$$

$$\nu'' = \nu'.$$

Finally, by transforming back to the system at rest K, we get for the reflected light

$$A''' = A''\,\frac{1 + \frac{v}{V}\cos\varphi''}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} = A\,\frac{1 - 2\frac{v}{V}\cos\varphi + \left(\frac{v}{V}\right)^2}{1 - \left(\frac{v}{V}\right)^2}\,,$$

$$\cos\varphi''' = \frac{\cos\varphi'' + \frac{v}{V}}{1 + \frac{v}{V}\cos\varphi''} = -\,\frac{\left[1 + \left(\frac{v}{V}\right)^2\right]\cos\varphi - 2\frac{v}{V}}{1 - 2\frac{v}{V}\cos\varphi + \left(\frac{v}{V}\right)^2}\,,$$

$$\nu''' = \nu''\,\frac{1 + \frac{v}{V}\cos\varphi''}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} = \nu\,\frac{1 - 2\frac{v}{V}\cos\varphi + \left(\frac{v}{V}\right)^2}{1 - \left(\frac{v}{V}\right)^2}\,. \qquad [33]$$

The energy (measured in the system at rest) striking the unit surface of the mirror per unit time is obviously $A^2/8\pi(V\cos\varphi - v)$. The energy leaving the unit surface of the mirror per unit time is $A'''^2/8\pi(-V\cos\varphi''' + v)$. According to the energy principle, the difference of these two expressions is the work done by the light pressure per unit time. Equating this work with $P\cdot v$, where P is the pressure of light, we obtain

$$P = 2\,\frac{A^2}{8\pi}\,\frac{\left[\cos\varphi - \frac{v}{V}\right]^2}{1 - \left[\frac{v}{V}\right]^2}\,. \qquad [34]$$

In first approximation, in agreement with experience and with other theories, we get [35]

$$P = 2\,\frac{A^2}{8\pi}\cos^2\varphi\,.$$

All problems in the optics of moving bodies can be solved by the method employed here. The essential point is that the electric and magnetic forces of light, which is influenced by a moving body, are transformed to a coordinate system that is at rest relative to that body. This reduces every problem in the optics of moving bodies to a series of problems in the optics of bodies at rest.

§9. *Transformation of the Maxwell-Hertz equations when convection currents are taken into consideration*

We start from the equations

$$\frac{1}{V}\left\{u_x\rho + \frac{\partial X}{\partial t}\right\} = \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z}\,, \qquad \frac{1}{V}\frac{\partial L}{\partial t} = \frac{\partial Y}{\partial z} - \frac{\partial Z}{\partial y}\,,$$

$$\frac{1}{V}\left\{u_y\rho + \frac{\partial Y}{\partial t}\right\} = \frac{\partial L}{\partial z} - \frac{\partial N}{\partial x}\,, \qquad \frac{1}{V}\frac{\partial M}{\partial t} = \frac{\partial Z}{\partial x} - \frac{\partial X}{\partial z}\,,$$

$$\frac{1}{V}\left\{u_z\rho + \frac{\partial Z}{\partial t}\right\} = \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}\,, \qquad \frac{1}{V}\frac{\partial N}{\partial t} = \frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x}\,,$$

where

$$\rho = \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}$$

denotes the 4π-fold density of electricity and (u_x, u_y, u_z) the electricity's velocity vector. If the electric masses are conceived as permanently bound to [36] small, rigid bodies (ions, electrons), then these equations constitute the electromagnetic foundation of Lorentz's electrodynamics and optics of moving bodies.

If, using the transformation equations presented in §3 and §6, we transform these equations, which should be valid in system K, to system k, we get the equations

$$\frac{1}{V}\left\{u_\xi \rho' + \frac{\partial X'}{\partial \tau}\right\} = \frac{\partial N'}{\partial \eta} - \frac{\partial M'}{\partial \zeta}, \qquad \frac{1}{V}\frac{\partial L'}{\partial \tau} = \frac{\partial Y'}{\partial \zeta} - \frac{\partial Z'}{\partial \eta},$$

$$\frac{1}{V}\left\{u_\eta \rho' + \frac{\partial Y'}{\partial \tau}\right\} = \frac{\partial L'}{\partial \zeta} - \frac{\partial N'}{\partial \xi}, \qquad \frac{1}{V}\frac{\partial M'}{\partial \tau} = \frac{\partial Z'}{\partial \xi} - \frac{\partial X'}{\partial \zeta},$$

$$\frac{1}{V}\left\{u_\zeta \rho' + \frac{\partial Z'}{\partial \tau}\right\} = \frac{\partial M'}{\partial \xi} - \frac{\partial L'}{\partial \eta}, \qquad \frac{1}{V}\frac{\partial N'}{\partial \tau} = \frac{\partial X'}{\partial \eta} - \frac{\partial Y'}{\partial \xi},$$

where

$$\frac{u_x - v}{1 - \frac{u_x v}{V^2}} = u_\xi ,$$

$$\frac{u_y}{\beta\left[1 - \frac{u_x v}{V^2}\right]} = u_\eta , \qquad \rho' = \frac{\partial X'}{\partial \xi} + \frac{\partial Y'}{\partial \eta} + \frac{\partial Z'}{\partial \zeta} = \beta\left[1 - \frac{v u_x}{V^2}\right]\rho .$$

$$\frac{u_z}{\beta\left[1 - \frac{u_x v}{V^2}\right]} = u_\zeta .$$

Since—as follows from the addition theorem of velocities (§5)—the vector (u_ξ, u_η, u_ζ) is actually the velocity of the electric masses measured in the system k, we have thus demonstrated that with our kinematic principles taken as a basis, the electrodynamic foundation of Lorentz's theory of the electrodynamics of moving bodies agrees with the principle of relativity.

Let me also briefly add that the following important proposition can easily be deduced from the equations we have derived: If an electrically charged body moves arbitrarily in space without change of its charge, observed from a coordinate system moving with the body, then its charge will also remain constant when observed from the system "at rest" K. [37]

§10. *Dynamics of the (slowly accelerated) electron*

In an electromagnetic field let a pointlike particle endowed with an electric charge ϵ (called "electron" in what follows) be in motion; about its law of motion we assume only the following:

If the electron is at rest during a particular epoch, its motion in the next element of time will occur according to the equations

$$\mu \frac{d^2x}{dt^2} = \epsilon X,$$

$$\mu \frac{d^2y}{dt^2} = \epsilon Y,$$

$$\mu \frac{d^2z}{dt^2} = \epsilon Z,$$

where x,y,z denote the coordinates of the electron and μ its mass, as long as the electron moves slowly.

Further, let the electron's velocity in some given time epoch be v. We seek to find the law by which the electron is moving in the next element of time.

Without affecting the generality of the consideration, we can and will assume that at the moment when we focus on it, the electron is at the coordinate origin, and is moving with velocity v along the X-axis of the coordinate system K. It is then obvious that at the instant indicated ($t = 0$), the electron is at rest relative to the coordinate system k that moves with constant velocity v parallel to the X-axis. [38]

From the above assumption combined with the relativity principle it is clear that, viewed from the system k, the electron will move during the immediately following time (for small values of t) according to the equations

$$\mu \frac{d^2\xi}{d\tau^2} = \epsilon X',$$

$$\mu \frac{d^2\eta}{d\tau^2} = \epsilon Y'$$

$$\mu \frac{d^2\zeta}{d\tau^2} = \epsilon Z',$$

where the symbols ξ, η, ζ, τ, X', Y', Z' refer to the system k. If we also stipulate that for $t = x = y = z = 0$ we should have $\tau = \xi = \eta = \zeta = 0$, then the transformation equations of §§3 and 6 will be valid, so that we get

$$\tau = \beta\left[t - \frac{v}{V^2} x\right],$$

$$\xi = \beta(x - vt), \qquad X' = X,$$

$$\eta = y, \qquad Y' = \beta\left[Y - \frac{v}{V} N\right],$$

$$\zeta = z, \qquad Z' = \beta\left[Z + \frac{v}{V} M\right].$$

With the help of these equations we transform the above equations of [39] motion from system k to system K and obtain

$$\frac{d^2x}{dt^2} = \frac{\epsilon}{\mu}\frac{1}{\beta^3} X,$$

$$\text{(A)} \qquad \frac{d^2y}{dt^2} = \frac{\epsilon}{\mu}\frac{1}{\beta}\left[Y - \frac{v}{V} N\right],$$

$$\frac{d^2z}{dt^2} = \frac{\epsilon}{\mu}\frac{1}{\beta}\left[Z + \frac{v}{V} M\right].$$

Following the usual approach, we now seek to determine the "longitudinal" and "transverse" masses of the moving electron. We write the [40] equations (A) in the form

$$\mu\beta^3 \frac{d^2x}{dt^2} = \epsilon X = \epsilon X',$$

$$\mu\beta^2 \frac{d^2y}{dt^2} = \epsilon\beta \left[Y - \frac{v}{V} N\right] = \epsilon Y',$$

$$\mu\beta^2 \frac{d^2z}{dt^2} = \epsilon\beta \left[Z + \frac{v}{V} M\right] = \epsilon Z',$$

and note first that $\epsilon X'$, $\epsilon Y'$, $\epsilon Z'$ are the components of the ponderomotive force exerted on the electron, as observed in a system co-moving at this instant with the electron at the latter's speed. (This force could be measured, for example, by a spring balance at rest in the last-mentioned system.) If we simply call this force "the force exerted on the electron," and maintain the equation [41]

Numerical value of mass × numerical value of acceleration = numerical value of force,

stipulating, in addition, that the accelerations be measured in the system at rest K, we obtain from the above equations

$$\text{Longitudinal mass} = \frac{\mu}{\left[\sqrt{1 - \left(\frac{v}{V}\right)^2}\right]^3},$$

$$\text{Transverse mass} = \frac{\mu}{1 - \left(\frac{v}{V}\right)^2}.$$

Of course, with a different definition of force and acceleration we would obtain different numerical values for the masses; this shows that we must proceed with great caution when comparing different theories of the [42] motion of the electron.

It should be noted that these results concerning mass are also valid for ponderable material points, since a ponderable material point can be made into an electron (in our sense) by adding to it an *arbitrarily small* electric charge.

We now determine the kinetic energy of the electron. If an electron starts out from the origin of the system K with an initial velocity 0 and is moving continually along the X-axis under the influence of an electrostatic force X, then it is clear that the energy drawn from the electrostatic field has the value $\int \epsilon X dx$. Since the electron is supposed to accelerate slowly and will therefore emit no energy in the form of radiation, the energy taken from the electrostatic field must be equated with the energy of motion W of the electron. Bearing in mind that the first of equations (A) holds during the entire process of motion considered, we obtain therefore

[43] $$W = \int \epsilon X dx = \int_0^v \beta^3 v dv = \mu V^2 \left[\frac{1}{\sqrt{1 - \left[\frac{v}{V}\right]^2}} - 1\right].$$

Thus, W becomes infinitely large when $v = V$. As in our previous
[44] results, superluminary velocities have no possibility of existence.

This expression for kinetic energy too must be valid for ponderable masses as well by virtue of the argument presented above.

Let us now enumerate those properties of the motion of the electron that result from the system of equations (A) and are accessible to experiment.

1. It follows from the second equation of the system of equations (A) that an electric force Y and a magnetic force N have an equally strong deflective effect on an electron moving with velocity v if $Y = N.v/V$. Thus we see that according to our theory we can determine the velocity of the
[45] electron for any arbitrary velocity from the ratio of the magnetic deflection A_m to the electric deflection A_e by applying the law

$$\frac{A_m}{A_e} = \frac{v}{V}.$$

This relation can be checked experimentally since the velocity of the electron can also be measured directly, e.g., using rapidly oscillating electric and magnetic fields.

2. It follows from the derivation for the kinetic energy of the electron that the potential difference traversed by the electron and the velocity v attained by it must be related by the equation

$$P = \int X dx = \frac{\mu}{\epsilon} V^2 \left[\frac{1}{\sqrt{1 - \left[\frac{v}{V}\right]^2}} - 1 \right].$$

3. We calculate the radius of curvature R of the path when a magnetic force N, which acts perpendicular to the velocity of the electron, is present (as the only deflecting force). From the second of equations (A) we obtain

$$-\frac{d^2y}{dt^2} = \frac{v^2}{R} = \frac{\epsilon}{\mu} \frac{v}{V} N . \sqrt{1 - \left[\frac{v}{V}\right]^2}$$

or

$$R = V^2 \frac{\mu}{\epsilon} \cdot \frac{\frac{v}{V}}{\sqrt{1 - \left[\frac{v}{V}\right]^2}} \cdot \frac{1}{N} .$$

These three relations are a complete expression of the laws by which the electron must move according to the theory presented here. [46]

In conclusion, let me note that my friend and colleague M. Besso steadfastly stood by me in my work on the problem here discussed, and that I am indebted to him for many a valuable suggestion. [47]

Bern, June 1905. (Received on 30 June 1905)

Doc. 24

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

by A. Einstein

[*Annalen der Physik* 18 (1905): 639-641]

The results of an electrodynamic investigation published by me recently in this journal[1] lead to a very interesting conclusion, which shall be derived here.

There I based myself upon the Maxwell-Hertz equations for empty space along with Maxwell's expression for the electromagnetic energy of space, and also on the following principle:

The laws governing the changes of state of physical systems do not depend on which one of two coordinate systems moving in uniform parallel translation relative to each other these changes of state are referred to (principle of relativity).

Based on these fundamental principles[2], I derived the following result, among others (*loc. cit.*, §8):

Let a system of plane waves of light, referred to the coordinate system (x,y,z), possess the energy ℓ; let the direction of the ray (the wave normal) form the angle φ with the x-axis of the system. If we introduce a new coordinate system (ξ,η,ζ), which is uniformly parallel-translated with respect to the system (x,y,z), and whose origin is moving along the x-axis with velocity v, then the above-mentioned quantity of light—measured in the system (ξ,η,ζ)—possesses the energy

$$\ell^* = \ell \frac{1 - \frac{v}{V}\cos\varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}},$$

where V denotes the velocity of light. We will make use of this result in the following.

---

[1] [1] A. Einstein, *Ann. d. Phys.* 17 (1905): 891.

[2] The principle of the constancy of the velocity of light used there is of course contained in Maxwell's equations.

Let there be a body at rest in the system (x,y,z), whose energy, referred to the system (x,y,z), is E_0. The energy of the body with respect to the system (ξ,η,ζ), which is moving with velocity v as above, shall be H_0.

Let this body simultaneously emit plane waves of light of energy $L/2$ (measured relative to (x,y,z)) in a direction forming an angle φ with the x-axis and an equal amount of light in the opposite direction. All the while, the body shall stay at rest with respect to the system (x,y,z). This process must satisfy the energy principle, and this must be true (according to the principle of relativity) with respect to both coordinate systems. If E_1 and H_1 denote the energy of the body after the emission of light, as measured relative to the system (x,y,z) and (ξ,η,ζ), respectively, we obtain, using the relation indicated above,

$$E_0 = E_1 + \left[\frac{L}{2} + \frac{L}{2}\right],$$

$$H_0 = H_1 + \left[\frac{L}{2}\,\frac{1 - \frac{v}{V}\cos\varphi}{\sqrt{1 - \left[\frac{v}{V}\right]^2}} + \frac{L}{2}\,\frac{1 + \frac{v}{V}\cos\varphi}{\sqrt{1 - \left[\frac{v}{V}\right]^2}}\right] = H_1 + \frac{L}{\sqrt{1 - \left[\frac{v}{V}\right]^2}} .$$

Subtracting, we get from these equations

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

The two differences of the form $H - E$ occurring in this expression have a simple physical meaning. H and E are the energy values of the same body, referred to two coordinate systems in relative motion, the body being at rest in one of the systems (system (x,y,z)). Hence it is clear that the difference $H - E$ can differ from the body's kinetic energy K with respect to the other system (system (ξ,η,ζ)) solely by an additive constant C, which depends on the choice of the arbitrary additive constants of the energies H and E. We can therefore put

$$H_0 - E_0 = K_0 + C$$

$$H_1 - E_1 = K_1 + C,$$

since C does not change during the emission of light. Thus, we get

$$K_0 - K_1 = L\left\{\frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} - 1\right\}.$$

The kinetic energy of the body with respect to (ξ,η,ζ) decreases as a result of the emission of light by an amount that is independent of the body's characteristics. Furthermore, the difference $K_0 - K_1$ depends on the velocity exactly like the kinetic energy of the electron (*loc. cit.*, §10).

Neglecting quantities of the fourth and higher orders, we can put

[2] $$K_0 - K_1 = \frac{L}{V^2}\frac{v^2}{2}.$$

From this equation it follows directly:

If a body releases the energy L in the form of radiation, its mass decreases by L/V^2. Since obviously here it is inessential that the energy withdrawn from the body happens to turn into energy of radiation rather than into some other kind of energy, we are led to the more general conclusion:

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 $L/9\cdot10^{20}$, if the energy is measured in ergs and the mass in grams. [3]

Perhaps it will prove possible to test this theory using bodies whose energy content is variable to a high degree (e.g., salts of radium).

If the theory agrees with the facts, then radiation transmits inertia between emitting and absorbing bodies. [4]

Bern, September 1905. (Received on 27 September 1905)

Doc. 25

Review of H. BIRVEN, *Fundamentals of the Mechanical Theory of Heat* (*Grundzüge der mechanischen Wärmetheorie.* Stuttgart and Berlin: F. Grub, 1905. 128 pp. 2.80 marks)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 950]

This booklet contains a concise, elementary exposition of the thermodynamics of gases and vapors as well as its application in the theory of the steam engine and refrigerating engines. Even though the booklet displays some inaccuracy with respect to the fundamental definitions and expositions (cf., e.g., the definition of entropy, p. 50), it will probably stand in good stead [1] to many an engineering student facing his exam with fragmentary college notebooks. [2]

Doc. 26

Review of A. PONSOT, "Heat in the Displacement of the Equilibrium of a Capillary System" ("Chaleur dans le déplacement de l'équilibre d'un système capillaire," *Académie des Sciences (Paris). Comptes rendus* 140 (1905): 1176-1179)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 952]

The author examines the foundations of the thermodynamic theory of capillarity and finds in them a not exactly correct assumption; nothing is said about the order of magnitude of the inaccuracies that arise from this [1] assumption.

Doc. 27

Review of K. BOHLIN, "On Impact Considered as the Basis of the Kinetic Theories of Gas Pressure and of Universal Gravitation" ("Sur le choc, considéré comme fondement des théories cinétiques de la pression des gaz et de la gravitation universelle," *Arkiv för Matematik, Astronomi och Fysik* 1 (1904): 522-540)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 952]

Starting from the remark that repulsive forces between particles are introduced into the kinetic theory of gases as well as into the dynamic theory [1] of gravitation for the sole purpose of explaining collisions, the author seeks [2] to avoid the introduction of repulsive forces altogether. He tries to ascribe the impact exclusively to the action of attractive forces between the corpuscles that constitute the colliding bodies. In doing so, he takes the position that every attractive force is to be explained (kinetically) by the effect of the impact of relatively infinitesimally small corpuscles, and every impact by the attraction of relatively infinitesimally small corpuscles. Thus, corpuscles of infinitely many orders of magnitudes are introduced to explain the elementary properties of matter.

Doc. 28

Review of G. MESLIN, "On the Constant in Mariotte and Gay-Lussac's Law" ("Sur la constante de la loi de Mariotte et Gay-Lussac," *Journal de physique théorique et appliquée* 4 (Series 4) (1905): 252-256)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 1114]

It is shown that the quotient of the above constant and the mechanical [1] equivalent of heat, which has a value slightly different from 2, is independent of the choice of the units for mass, length, and time, but that nevertheless no physical meaning is to be ascribed to the numerical value of this quotient because it depends on the choice of the unit of molecular weight. [2]

Doc. 29

Review of A. FLIEGNER, "The Efflux of Hot Water from Container Orifices" ("Das Ausströmen heissen Wassers aus Gefässmündungen," *Schweizerische Bauzeitung* 45 (1905): 282-285, 306-308)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 1115]

According to Zeuner and Lorenz, the deviation of Pulin and Bonnin's [1] experimental results concerning the efflux velocity of water under the [2] pressure of its steam is to be ascribed to a sort of evaporation delay. In contrast, the author takes the point of view that the lack of agreement between the experiments and the theory is to be ascribed to the fact that in the above-mentioned experiments, due to temperature differences within the water container, the water reaching the point of outflow has a somewhat lower temperature than that corresponding to the vapor pressure in the container. In addition, an experiment is cited in which the above source of error was avoided and in which the result was in agreement with the theory.

Doc. 30

Review of J. J. WEYRAUCH, *An Outline of the Theory of Heat. With Numerous Examples and Applications. Part 1* (*Grundriss der Wärmetheorie. Mit zahlreichen Beispielen und Anwendungen.* Stuttgart: K. Wittwer, 1905. 131 pp.)

[*Beiblätter zu den Annalen der Physik* 29 (1905): 1152]

This book is based on lectures the author gave at the Stuttgart Technical University and contains in the main the theory of the basic laws of thermodynamics and, subsequent to that, in a clear and comprehensive presentation, the theories of the various heat engines. The book is very well suited for private study since much care has been devoted to the didactic aspects in order to accomplish the above purpose. In order to present the theory in a compact and clear form, as well as to impress the abstract results obtained upon the reader's mind as vividly as possible, numerous examples and problems are inserted between the theoretical presentations, which are by no means restricted to applications important to the engineer. Many examples are taken from the history of the theory of heat, especially the trains of thought of Robert Mayer are presented in detail; the workings and ways of functioning
[1] of "human engines" are also subjected to a detailed consideration and compared with artificial heat engines. The seven sections of the book's first half under consideration are titled as follows: I. Conservation of energy. The first law. II. Heat and work. The second law. III. On heat engines in general. IV. On gases. V. On air engines. VI. On chemistry and the kinetic theory of gases. VII. On combustion engines.— No previous knowledge of engineering and physics is assumed, but elements of infinitesimal calculus are. Constructional details of heat engines are entered into only insofar as this is required for the presentation of the theories of the different engine types and for the study of their efficiency. A table of symbols used, and a name and subject index, are appended to the volume, so that the book can also be profitably used as a reference work, and each of the many experimental values given (e.g., caloric values of different fuels, efficiency of various heat engines achieved in practice to date, etc.,) can easily be

found. As mentioned in the introduction, the second half of the book is expected to appear during the next year. It will contain the theory of saturated and superheated vapors, sections on aerostatics, aerodynamics, and solid bodies, as well as the corresponding applications, in a presentation very similar to that in the first half. [2]

Doc. 31

Review of A. FLIEGNER, "On the Thermal Value of Chemical Processes"
("Über den Wärmewert chemischer Vorgänge," *Naturforschende Gesellschaft in Zürich. Vierteljahrsschrift* 50 (1905): 201-212)
[*Beiblätter zu den Annalen der Physik* 29 (1905): 1158]

The author overlooks the fact that the defining equation of entropy $dS = dQ/T$ holds for *reversible* processes only, and as a consequence he arrives at the result that one cannot speak of a change of a system's entropy by a chemical process.

Doc. 32

ON THE THEORY OF BROWNIAN MOTION

by A. Einstein

[*Annalen der Physik* 19 (1906): 371-381]

Soon after the publication of my paper on the motion of particles
[2] suspended in liquids demanded by the molecular theory of heat,[1] Mr. Siedentopf
[3] (Jena) informed me that he and other physicists—Prof. Gouy (Lyon) probably having been the first—had become convinced by direct observation that the so-called Brownian motion is caused by the random thermal motion of the liquid's molecules.[2] Not only the qualitative properties of Brownian motion but also the order of magnitude of the paths traversed by the particles are in full agreement with the results of the theory. I shall not compare here the meager experimental material available to me with the results of the theory, but shall leave this comparison to those engaged in experimental investigation
[5] of this topic.

The present paper shall supplement my above-mentioned paper in several points. We will derive here not only the translatory, but also the rotational motion of suspended particles for the simplest special case when the particles
[6] have a spherical shape. We will also establish the shortest observation times for which the result given in the paper is still valid.

We will use here a more general method of derivation, partly to show how Brownian motion relates to the foundations of the molecular theory of heat, and partly to be able to derive the formulas for the translatory and for the rotational motion by a common investigation. Let us assume that α is an observable parameter of a physical system in thermal equilibrium and that the system is in so-called indifferent equilibrium at every (possible) value of α. According to classical thermodynamics, which makes a *fundamental* distinction between heat and other kinds of energy, spontaneous changes of α do not take place, but according to the molecular theory of heat they do. In the following we will investigate what laws these changes must obey according to the

---

[1] [1] A. Einstein, *Ann. d. Phys.* 17 (1905): 549.

[4] [2] M. Gouy, *Jour. de Phys.* 7, No. 2 (1888): 561.

latter theory. We will then have to apply these laws to the following special cases:

1. α is the x-coordinate of the center of gravity of a spherically shaped particle suspended in a homogeneous liquid (which is not subject to gravitation).

2. α is the angle of rotation that determines the position of a spherical particle suspended in a liquid and capable of rotating about a diameter.

§*1*. *On a case of thermodynamic equilibrium*

In an environment of absolute temperature T let there be a physical system in thermal interaction with this environment and in a state of thermal equilibrium. This system, which hence also possesses the absolute temperature T, shall be completely determined by the state variables $p_1 \ldots p_n$ according to the molecular theory of heat.[1] In the special cases to be considered, we can choose for the state variables $p_1 \ldots p_n$ the coordinates and velocity components of all atoms constituting the system under consideration.

The probability that at a randomly chosen instant of time all state variables $p_1 \ldots p_n$ will lie in the n-fold infinitesimally small region $(dp_1 \ldots dp_n)$ is given by the equation[2]

$$dw = Ce^{-\frac{N}{RT}E}\, dp_1 \ldots dp_n, \tag{1}$$

where C denotes a constant, R the universal constant of the gas equation, N the number of true molecules per gram-molecule, and E the energy.

Suppose that α is an observable parameter of the system and that to each system of values $p_1 \ldots p_n$ there corresponds a definite value α. We denote by $A d\alpha$ the probability that at a randomly chosen instant the value of the parameter α will lie between α and $\alpha + d\alpha$. We then have

---

[1] Cf. *Ann. d. Phys.* 17 (1905): 549. [7]

[2] *Loc. cit.*, §§3 and 4. [8]

$$A d\alpha = \int_{d\alpha} C e^{-\frac{N}{RT} E} dp_1 \ldots dp_n, \tag{2}$$

where the integral on the right-hand side is extended over all combinations of those state variables whose value of α lies between α and $\alpha + d\alpha$.

We will confine ourselves to the case in which the nature of the problem makes it immediately evident that all (possible) values of α have the same [9] probability (frequency), i.e., that the quantity A is independent of α.

Imagine now a second physical system that differs from the system just considered by the sole fact that it is acted upon by a force of potential $\Phi(\alpha)$, which depends only on α. If E is the energy of the system considered earlier, then $E + \Phi$ will be the energy of the system considered now, so that we get the following relation, analogous to equation (1):

$$dw' = C' e^{-\frac{N}{RT}(E+\Phi)} dp_1 \ldots dp_n.$$

This, in turn, yields a relation analogous to equation (2) for the probability dW that at an arbitrarily chosen instant the value of α will lie between α and $\alpha + d\alpha$:

$$\begin{aligned} dW = \int C' e^{-\frac{N}{RT}(E+\Phi)} dp_1 \ldots dp_n &= \frac{C'}{C} e^{-\frac{N}{RT}\Phi} A d\alpha \\ &= A' e^{-\frac{N}{RT}\Phi} d\alpha, \end{aligned} \tag{I}$$

where A' is independent of α.

This relation, which corresponds exactly to the exponential law used frequently by Boltzmann in his investigations on the theory of gases, is [10] characteristic for the molecular theory of heat. It determines how much a parameter of a system subjected to a constant external force diverges from the value corresponding to stable equilibrium because of random molecular motion.

§2. *Examples of application of the equation derived in §1*

We consider a body whose center of gravity can move along a straight line (the X-axis of a coordinate system). The body shall be surrounded by a gas, and there shall be thermal and mechanical equilibrium. According to the molecular theory, the body will move back and forth along the straight line in a random fashion due to the nonuniformity of molecular collisions, such that none of the points of the straight line will be preferred in this motion — provided that no forces other than those of molecular collision are exerted on the body in the direction of the straight line. Hence, the abscissa x of the center of gravity is a parameter of the system, which possesses the properties stipulated above for the parameter α.

We will now introduce a force $K = -Mx$ that acts on the body in the direction of the straight line. According to the molecular theory the center of gravity of the body will then also carry out random motions, but without deviating too far from the point $x = 0$, whereas according to classical thermodynamics it must be at rest at the point $x = 0$. According to the molecular theory (formula I),

$$dW = A'e^{-\frac{N}{RT}M\frac{x^2}{2}}\,dx$$

equals the probability that at a randomly chosen instant the value of the abscissa lies between x and $x + dx$. From this we find the mean distance of the center of gravity from the point $x = 0$,

$$\sqrt{\overline{x^2}} = \frac{\int_{-\infty}^{+\infty} x^2 A'e^{-\frac{N}{RT}\frac{Mx^2}{2}}\,dx}{\int_{-\infty}^{+\infty} A'e^{-\frac{N}{RT}\frac{Mx^2}{2}}\,dx} = \sqrt{\frac{RT}{NM}}\,. \qquad [11]$$

For $\sqrt{\overline{x^2}}$ to be large enough to be accessible to observation, the force that determines the body's equilibrium position must be very small. Putting $\sqrt{\overline{x^2}} = 10^{-4}$ cm as the lower limit of observability, we get $M =$ about $5\cdot10^{-6}$ [12]

for $T = 300$. Thus, for the body to perform fluctuations observable under the microscope, the force acting on it must not exceed five millionths of a dyne
[13] for an elongation of 1 cm.

Let us add one further theoretical remark to the equation derived.
[14] Suppose the body under consideration carries an electric charge distributed over a very small space, and the gas surrounding the body is so rarefied that the body performs sinusoidal oscillations only slightly modified by the surrounding gas. The body then radiates electric waves into space and absorbs energy from the radiation of the surrounding space; it thus mediates an exchange of energy between radiation and gas. We can derive the limiting law of thermal radiation, which seems to hold for long wave lengths and high temperatures, by formulating the condition that the body in question emits on the average as much radiation as it absorbs. We arrive in this way[1] at the following formula for the radiation density ρ_ν that corresponds to the frequency ν:

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

where L denotes the velocity of light.

The radiation formula given by Mr. Planck[2] reduces to this formula at low frequencies and high temperatures. From the coefficient of the limiting law we can determine the quantity N and thus arrive at Planck's determina-
[17] tion of the elementary quanta. The fact that in the way indicated we do not obtain the true law of radiation, but only a limiting law, seems to me to be
[18] rooted in a fundamental imperfection of our physical conceptions.

We will also use formula (I) to decide how small the suspended particle needs to be to remain permanently suspended despite the effect of gravity. We can confine ourselves to the case that the particle has a greater specific gravity than the liquid, since the opposite case is completely analogous.

If v is the volume of the particle, ρ its density, ρ_0 the density of the liquid, g the acceleration of gravity, and x the vertical distance of a point from the bottom of the container, equation (I) will yield

---

[15] [1] Cf. *Ann. d. Phys.* 17 (1905): 549, §§1 and 2.

[16] [2] M. Planck, *Ann. d. Phys.* 1 (1900): 99.

$$dW = \text{const.}\, e^{-\frac{N}{RT} v(\rho - \rho_0) g x} dx.$$

Thus we will then find that suspended particles can float in a liquid if, for values of x that do not escape observation because of their smallness, the quantity

$$\frac{N}{RT} v(\rho - \rho_0) g x$$

does not have too high a value—provided that the particles that have reached the bottom of the container do not adhere there due to some circumstance or other. [19]

§3. *On the changes in the parameter α caused by thermal motion* [20]

Now we return again to the general case discussed in §1, for which we derived equation (I). For the sake of a simpler mode of expression and visualization, we will now assume, however, that a very large number (n) of identical systems of the kind described there are involved; in that case we have to deal with numbers instead of probabilities. Equation (I) expresses then the following:

Of N systems, there are

$$dn = \varphi e^{-\frac{N}{RT}\Phi} d\alpha = F(\alpha) d\alpha \tag{Ia}$$

systems in which the value of the parameter α lies between α and $\alpha + d\alpha$ at a randomly chosen instant.

We shall use this relation to determine the magnitude of the irregular changes of the parameter α produced by the random thermal processes. To that end, we express in symbols that, within the time span t, the function $F(\alpha)$ does not change under the combined effect of the force corresponding to the potential Φ and the random thermal process; here t denotes a time so small that the corresponding changes of the quantities α of the individual systems can be considered as infinitesimally small changes in the argument of the function $F(\alpha)$.

If lengths numerically equal to α are plotted along a straight line starting from some specified origin, then to each system there will correspond a point (α) on this straight line. $F(\alpha)$ is the density of the system-points (α) on the line. During time t, exactly as many system-points must then cross an arbitrary point (α_0) of the line in one direction as in the opposite one.

Let a force corresponding to the potential Φ produce a change of magnitude

$$\Delta_1 = -B\frac{\partial\Phi}{\partial\alpha}t$$

in α, where B is independent of α, i.e., the velocity of change of α shall be proportional to the operating force and independent of the value of the parameter. We will call the factor B "the mobility of the system with respect to α."

Thus, if the external force were to operate without the quantity α being changed by the random molecular thermal process, then

$$n_1 = B\left[\frac{\partial\Phi}{\partial\alpha}\right]_{\alpha=\alpha_0}.t.F(\alpha_0)$$

system-points would cross the point (α_0) toward the negative side during time t.

Let $\psi(\Delta)$ be the probability that, due to the random thermal process, the parameter α of a system experiences during time t a change whose value lies between Δ and $\Delta + d\Delta$, where $\psi(\Delta) = \psi(-\Delta)$, and ψ is independent of α. The number of system-points crossing the point (α_0) toward the positive side on account of the random thermal process during time t is then

$$n_2 = \int_{\Delta=0}^{\Delta=\infty} F(\alpha_0 - \Delta)\chi(\Delta)\,d\Delta,$$

where we have put

$$\int_{\Delta}^{\infty} \psi(\Delta)\,d\Delta = \chi(\Delta).$$

The number of system-points traveling toward the negative side on account of the random thermal process is

$$n_3 = \int_{\Delta}^{\infty} F(\alpha_0 + \Delta)\chi(\Delta)\,d\Delta.$$ [21]

The mathematical expression for the invariability of the function F is thus

$$-n_1 + n_2 - n_3 = 0.$$

If we substitute the expressions found for n_1, n_2, n_3, and take into account that Δ is infinitesimally small, and that $\psi(\Delta)$ differs from zero only for infinitesimal values of Δ, we obtain after simple calculation

$$B\left[\frac{\partial\Phi}{\partial\alpha}\right]_{\alpha=\alpha_0} F(\alpha_0)\,t + \frac{1}{2}\,F'(\alpha_0)\overline{\Delta^2} = 0.$$

Here

$$\overline{\Delta^2} = \int_{-\infty}^{+\infty} \Delta^2\psi(\Delta)\,d\Delta$$

denotes the mean of the squares of the changes of the quantities α produced by the irregular thermal process during time t. From this relation we obtain, if we take into consideration equation (Ia),

(II) $$\sqrt{\overline{\Delta^2}} = \sqrt{\frac{2R}{N}} \cdot \sqrt{BTt}\ .$$

Here R denotes the constant of the gas equation ($8.31 \cdot 10^7$), N the number of true molecules in a gram-molecule (about $4 \cdot 10^{23}$), B the "mobility of the system with respect to the parameter α," T the absolute temperature, and t the time within which the changes in α produced by the random thermal process take place. [22]

§4. *Application of the derived equation to Brownian motion*

Using equations (II), we now calculate the mean displacement in a particular direction (the X-direction of a coordinate system) experienced during time t by a spherical body suspended in a liquid. To this end we must substitute the corresponding value for B in the above equation.

If a force is exerted on a sphere of radius P that is suspended in a liquid with a coefficient of friction k, the sphere will move with velocity[1] $K/6\pi kP$. Hence we have to put

$$B = \frac{1}{6\pi kP},$$

so that—in conformity with the paper cited above—for the mean displacement of the suspended sphere in the direction of the X-axis we obtain the value

$$\sqrt{\overline{\Delta_x^2}} = \sqrt{t}\sqrt{\frac{RT}{N}\frac{1}{3\pi kP}}.$$

Second, we consider the case when the sphere in question is pivoted in the liquid such that it can freely rotate (without bearing friction) about one of its diameters, and we seek to determine the mean rotation $\sqrt{\overline{\Delta_r^2}}$ of the sphere produced by the random thermal process during time t.

If a torque D acts upon a sphere of radius P that is pivoted in a liquid whose coefficient of friction is k, the sphere will rotate with the angular velocity[2]

[24] $$\psi = \frac{D}{8\pi kP^3}.$$

Accordingly, we have to put

$$B = \frac{1}{8\pi kP^3}.$$

We thus get

$$\sqrt{\overline{\Delta_r^2}} = \sqrt{t}\sqrt{\frac{RT}{N}\frac{1}{4\pi kP^3}}.$$

---

[23] [1] Cf. G. Kirchhoff, *Vorles. über Mechanik* [Lectures on Mechanics]. Lecture 26.

[2] *ibid.*

Thus, the rotational motion produced by molecular motion decreases much faster with increasing P than does translational motion.

For $P = 0.5$ mm and water at 17°, the formula yields about 11 seconds of arc for the angle traversed in one second on the average, and about 11 minutes of arc for that traversed in one hour. For $P = 0.5$ micron and water at 17°, we get about 100 degrees of arc for $t = 1$ sec. [25]

In the case of a freely floating suspended particle, three mutually independent rotational motions of this kind take place.

The formula derived for $\sqrt{\overline{\Delta^2}}$ might be applied to other cases as well. For example, if the reciprocal of the electric resistance of a closed circuit is substituted for B, the formula shows how much electricity will flow on the average through some particular cross section of the conductor during time t, which relation is connected again with the limiting law of black-body radiation for great wave lengths and high temperatures. However, since I could not [26] find any additional experimentally verifiable consequences, any treatment of further special cases seems useless to me.

§5. *On the limit of validity of the formula for* $\sqrt{\overline{\Delta^2}}$ [27]

It is clear that formula (II) cannot be valid for arbitrarily small time intervals. This is so because the mean velocity of the change of α resulting from the thermal process,

$$\frac{\sqrt{\overline{\Delta^2}}}{t} = \sqrt{\frac{2RTB}{N}} \cdot \frac{1}{\sqrt{t}},$$

becomes infinitely large for an infinitesimally small time interval t, which is obviously impossible because every suspended body would then have to move with infinitely great instantaneous velocity. The reason for this is that we have implicitly assumed in our derivation that the process occurring during time t is to be conceived as an event that is independent of the process occurring during the times immediately preceding it. But the shorter the times t chosen, the less this assumption applies. For if at time $z = 0$ the instantaneous value of the velocity of change were

[28]
$$\frac{d\alpha}{dt} = \beta_0 ,$$

and if in some subsequent time interval the velocity of change β were not influenced by the random thermal process but the change of β were determined by the passive resistance $(1/B)$ alone, $d\beta/dz$ would obey the relation

$$-\mu \frac{d\beta}{dz} = \frac{\beta}{B} .$$

μ is defined here by the stipulation that $\mu(\beta^2/2)$ should be the energy that corresponds to the velocity of change β. Thus, in the case of translational motion of a suspended sphere, e.g., $\mu(\beta^2/2)$ would be the kinetic energy of the sphere plus the kinetic energy of the co-moving liquid. Integrating, we get

$$\beta = \beta_0 e^{-\frac{z}{\mu B}} .$$

From this result one concludes that formula (II) holds only for time [29] intervals that are large compared with B.

For corpuscles with a diameter of 1 micron and density $\rho = 1$ in water at room temperature, the lower limit of validity of formula (II) is about 10^{-7} seconds; this lower limit for time intervals increases as the square of the corpuscle radius. Both facts hold true for the translational as well as the rotational motion of particles.

Bern, December 1905. (Received on 19 December 1905)

Doc. 33

"SUPPLEMENT" TO "A NEW DETERMINATION OF MOLECULAR DIMENSIONS"

[Appended to the version of Document 15 that was published as an article in *Annalen der Physik* 19 (1906): 289-305]

[*Annalen der Physik* 19 (1906): 305-306]

The new edition of Landolt and Börnstein's Physical-Chemical Tables [1] contains data that are much more useful for calculating the size of the sugar molecule and the number N of real molecules in a gram-molecule.

Thovert found (Tables, p. 372) that the coefficient of diffusion of [2] sugar in water at 18.5°C at a concentration of 0.005 mol/liter has the value of 0.33 cm^2/day. Further, from a table containing observed values obtained by Hosking (Tables, p. 81) we can find by interpolation that in a dilute sugar [3] solution a 1% increase in sugar content at 18.5°C corresponds to a 0.00025 increase in the coefficient of viscosity. [4]

Based on these data, one finds

$$P = 0.78 \cdot 10^{-6} \text{ mm}$$

and

$$N = 4.15 \cdot 10^{23}.$$

Bern, January 1906.

Doc. 34

ON THE THEORY OF LIGHT PRODUCTION AND LIGHT ABSORPTION

by A. Einstein

[*Annalen der Physik* 20 (1906): 199-206]

In a study published last year[1] I showed that the Maxwell theory of
[2] electricity in conjunction with the theory of electrons leads to results that contradict the evidence on black-body radiation. By a route described in that study, I was led to the view that light of frequency ν can only be absorbed or emitted in quanta of energy $(R/N)\beta\nu$, where R denotes the absolute constant of the gas equation applied to one gram-molecule, N the number of actual molecules in one gram-molecule, β the exponential coefficient of Wien's (and Planck's) radiation formula, and ν the frequency of the light in question. This relationship was developed for a range that corresponds to the
[3] range of validity of Wien's radiation formula.

At that time it seemed to me that in a certain respect Planck's theory of radiation[2] constituted a counterpart to my work. New considerations, which are being reported in §1 of this paper, showed me, however, that the theoretical foundation on which Mr. Planck's radiation theory is based differs from the one that would emerge from Maxwell's theory and the theory of electrons, precisely because Planck's theory makes implicit use of the aforementioned hypothesis of light quanta.

In §2 of this paper I shall make use of the hypothesis of light quanta to derive a relationship between the Volta effect and photoelectric diffusion.

§1. Planck's theory of radiation and the light quanta

In §1 of my paper cited above I have shown that the molecular theory of heat combined with the Maxwell theory of electricity and the theory of

---

[1] [1] A. Einstein, *Ann. d. Phys.* 17 (1905): 132.

[4] [2] M. Planck, *Ann. d. Phys.* 4 (1901): 561.

electrons lead to a formula for black-body radiation that contradicts experience

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

Here ρ_ν denotes the density of radiation at temperature T and at a frequency between ν and $\nu + 1$.

What is the reason that Mr. Planck did not arrive at the same formula, but obtained instead the expression

(2) $$\rho_\nu = \frac{\alpha\nu^3}{e^{\frac{\beta\nu}{T}} - 1} \quad ?$$ [5]

Mr. Planck derived[1] the mean energy $\bar{E}_\nu$ of a resonator of proper frequency ν situated in a space filled with disordered radiation as given by the equation [7]

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

This reduced the problem of black-body radiation to the problem of determining $\bar{E}_\nu$ as a function of temperature. The latter problem will have been solved if one can calculate the entropy of one of many similarly constituted, mutually interacting resonators of proper frequency ν that are in dynamic equilibrium.

Let us envision the resonators as ions that could perform rectilinear sinusoidal vibrations about an equilibrium position. The fact that the ions have electrical charges is irrelevant in the calculation of this entropy; we simply have to conceive these ions as mass points (atoms) whose momentary state is completely determined by their instantaneous deviation x from the equilibrium position and by their instantaneous velocity $dx/dt = \xi$.

---

[1] M. Planck, *Ann. d. Phys.* 1 (1900): 99. [6]

For the distribution of states of these resonators to be uniquely
determined in thermodynamic equilibrium, one has to assume that there exists
an arbitrarily small number of freely moving molecules besides the resonators,
which by virtue of their collisions with the ions can transmit energy from
[9] resonator to resonator; we will not take into account these molecules in our
calculation of entropy.

We could determine $\bar{E}_\nu$ as a function of temperature from the Maxwell-
[10] Boltzmann distribution law and would thereby obtain the invalid radiation
formula (1). One arrives at the route taken by Mr. Planck in the following
manner.

Let $p_1, \ldots p_n$ be appropriately chosen state variables[1] that completely determine the state of a physical system (e.g., in our case the values x and ξ of all the resonators). At the absolute temperature T, the entropy S of this system is represented by the equation[2]

[12] (4)
$$S = \frac{\bar{H}}{T} + \frac{R}{N} \lg \int e^{-\frac{N}{RT} H} \, dp_1 \ldots dp_n \,,$$

where $\bar{H}$ denotes the energy of the system at temperature T, H denotes the energy as a function of $p_1, \ldots p_n$, and the integral is to be extended over all possible combinations of the values of $p_1, \ldots p_n$.

If the system consists of a very large number of molecular structures--
and the formula has meaning and validity only in this case--then only those
[13] combinations of values of the $p_1 \ldots p_n$ whose H differs very little from $\bar{H}$
contribute significantly to the value of the integral appearing in S.[3] If
this is taken into account, it is easily seen that, except for negligible
quantities, one can put

[14]
$$S = \frac{R}{N} \lg \int_{H}^{H+\Delta H} dp_1 \ldots dp_n \,,$$

---

[11] [1] A. Einstein, *Ann. d. Phys.* 11 (1903): 170.

[2] *loc. cit.* §6.

[3] Follows from §3 and §4 *loc. cit.*

where ΔH should be chosen very small, yet large enough to make $R\ \lg(\Delta H)/N$ a negligible quantity. S is then independent of the value of ΔH.

If one substitutes the variables x_α and ξ_α of the resonators instead of $dp_1, \ldots dp_n$ in the equation and takes into account that the equation holding for the α-th resonator is

$$\int_{E_\alpha}^{E_\alpha + dE_\alpha} dx_\alpha d\xi_\alpha = \text{const. } dE_\alpha$$

(because E_α is a quadratic, homogeneous function of x_α and ξ_α), one obtains the following expression for S:

(5) $$S = \frac{R}{N} \lg W ,$$

where one has put

(5a) $$W = \int_H^{H+\Delta H} dE_1 \ldots dE_n .$$

If one would calculate S according to this formula, one would again arrive at the invalid radiation formula (1). To arrive at Planck's formula, [15] one has to postulate that, rather than assume any value whatsoever, the energy E_α of a resonator can only assume values that are integral multiples of ϵ, where

$$\epsilon = \frac{R}{N} \beta\nu .$$

This is because, on setting $\Delta H = \epsilon$, one immediately sees from equation (5a) that, except for an inconsequential factor, W turns into the very quantity that Mr. Planck named "the number of complexions." [16]

Hence, we must view the following proposition as the basis underlying Planck's theory of radiation:

The energy of an elementary resonator can only assume values that are integral multiples of $(R/N)\beta\nu$; by emission and absorption, the energy of a resonator changes by jumps of integral multiples of $(R/N)\beta\nu$.

However, this assumption involves yet a second one, because it contradicts the theoretical basis from which equation (3) is developed. For if the energy of a resonator can only change in jumps, then the mean energy of a resonator in a radiation space cannot be obtained from the usual theory of electricity, because the latter does not recognize *distinguished* energy values of a resonator. Thus, the following assumption underlies Planck's theory:

Although Maxwell's theory is not applicable to elementary resonators, nevertheless the *mean* energy of an elementary resonator in a radiation space
[17] is equal to the energy calculated by means of Maxwell's theory of electricity.

This proposition would be immediately plausible if, in all those parts of the spectrum that are relevant for observation, $\epsilon = (R/N)\beta\nu$ were small compared with the mean energy $\bar{E}_\nu$ of a resonator; however, this is not at all the case, for within the range of validity of Wien's radiation formula, $e^{\beta\nu/T}$ is large compared with 1. It is easy to prove that according to Planck's theory of radiation, within the range of validity of Wien's radiation formula, $\bar{E}_\nu/\epsilon$ has the value $e^{-\beta\nu/T}$, thus, $\bar{E}_\nu$ is much smaller than ϵ. Therefore only a few resonators have energies different from zero.

In my opinion the above considerations do not at all disprove Planck's theory of radiation; rather, they seem to me to show that with his theory of radiation Mr. Planck introduced into physics a new hypothetical element: the hypothesis of light quanta.

§2. *An expected quantitative relationship between photoelectric diffusion and the Volta effect*

It is well known that if metals are ordered according to their photoelectric sensitivity, one obtains the Volta electric potential series, in
[18] which a metal is the more photosensitive the closer it is to the electro-
[19] positive end of the electric potential series.

To a certain degree, this fact can be understood by assuming only that the forces (which are not to be examined here) that produce the active double layers reside on the metal-gas interface rather than on the metal-metal interface.

Let these forces produce an electric double layer on the surface of a piece of metal M that borders on a gas, and a corresponding potential difference V between metal and gas, taken as positive when the metal has the higher potential.

Let V_1 and V_2 be the potential differences between metals M_1 and M_2 in electrostatic equilibrium if they are insulated from each other. If the two metals are brought into contact, the electric equilibrium is disturbed and complete[1] voltage equalization of the metals takes place. Thereby, simple layers will be superposed on the aforementioned double layers at the metal-gas interfaces; to these corresponds an electrostatic field in the air space whose line integral equals the voltage difference.

If V_{ℓ_1} and V_{ℓ_2} denote the electric potentials at points of the gas space directly adjacent to the metals in contact, and V' denotes the potential in the interior of the metals, we have

$$V' - V_{\ell_1} = V_1 ,$$

$$V' - V_{\ell_2} = V_2 ,$$

and thus

$$V_{\ell_2} - V_{\ell_1} = V_1 - V_2 .$$

Thus, the electrostatically measurable Volta difference is numerically equal to the difference of the potentials assumed by the metals in the gas if they are insulated from each other.

If one ionizes the gas, the electric forces present in the gas space will cause a migration of the ions, to which there corresponds a current in the metals which, at the place of contact of the metals, is directed from the metal with the higher V (less electropositive) to the metal with the lower [20]
V (more electropositive). [21]

Suppose a metal M is insulated in a gas. Let V be its potential difference with respect to the gas that corresponds to the double layer. In order to move a unit of negative electricity from the metal into the gas, an amount of work numerically equal to the potential V has to be performed. Hence, the greater the V, i.e., the less electropositive the metal, the more [22]

---

[1] We disregard the effect of thermoelectric forces.

energy is needed for the photoelectric diffusion, i.e., the smaller the photoelectric sensitivity of the metal.

So far we have considered the facts without making assumptions about the nature of photoelectric diffusion. However, the hypothesis of light quanta also yields a quantitative relationship between the Volta effect and photoelectric diffusion. Thus, to move a negative elementary quantum (charge ϵ) from the metal into the gas, it has to be supplied with at least an energy $V\epsilon$. Then, a light species will be able to remove negative electricity from the metal only when the "light quantum" of that light species has at least the value $V\epsilon$. We thus obtain

$$V\epsilon \leqq \frac{R}{N}\beta\nu$$

or

$$V \leqq \frac{R}{A}\beta\nu ,$$

where A denotes the charge of one gram-molecule of a univalent ion.

If we now assume that some of the absorbing electrons are able to leave the metal as soon as the energy of the light quanta exceeds $V\epsilon$[1] — which is a very plausible assumption — we obtain

$$V = \frac{R}{A}\beta\nu ,$$

where ν denotes the lowest photoelectrically effective frequency.

Thus, if ν_1 and ν_2 are the lowest light frequencies acting on the metals M_1 and M_2, the following equation will hold for the Volta potential difference V_{12} of the two metals:

$$-V_{12} = V_1 - V_2 = \frac{R}{A}\beta(\nu_1 - \nu_2),$$

or, if V_{12} is measured in volts:

$$V_{12} = 4.2 \times 10^{-15}(\nu_2 - \nu_1) .$$

---

[1] The thermal energy of electrons is disregarded.

This formula contains the following, at least by and large valid, proposition: The more electropositive a metal, the smaller the lowest light frequency that is effective for that metal. It would be of great interest to [23] know whether this formula expresses the facts in a quantitative way as well. [24]

Bern, March 1906. (Received on 13 March 1906)

Doc. 35

THE PRINCIPLE OF CONSERVATION OF MOTION OF THE CENTER OF GRAVITY AND THE INERTIA OF ENERGY

by A. Einstein

[*Annalen der Physik* 20 (1906): 627-633]

In a paper published last year[1] I showed that Maxwell's electromagnetic equations in conjunction with the principle of relativity and the principle of energy conservation led to the conclusion that the mass of a body changes with the change in its energy content, no matter what kind of change of energy this may be. It turned out that to an energy change of magnitude ΔE there must correspond a change of mass of the same sign and of magnitude $\Delta E/V^2$, where V denotes the velocity of light.

In the present paper I want to show that the above theorem is the necessary and sufficient condition for the law of the conservation of motion of the center of gravity to be valid (at least in first approximation) also for systems in which not only mechanical, but also electromagnetic processes take place. Although the simple formal considerations that have to be carried out to prove this statement are in the main already contained in a work by H. Poincaré[2], for the sake of clarity I shall not base myself upon that work.

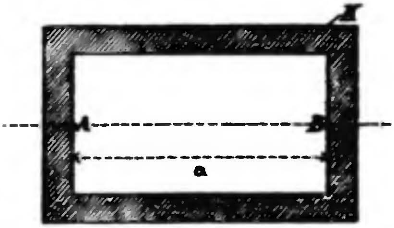
§*1.* *A special case*

Let K be a stationary rigid hollow cylinder freely floating in space. Let there be in A an arrangement for sending a certain amount S of radiating energy through the cavity to B. During the emission of this quantity of radiation a radiation pressure acts upon the left interior wall of the tube K, imparting to the latter a certain velocity that is directed to the left. If the hollow cylinder's mass is M, then this velocity equals $\frac{1}{V} \cdot \frac{S}{M}$, as can be proved easily from the laws of radiation pressure, where V denotes the

---

[1] [1] A. Einstein, *Ann. d. Phys.* 18 (1905): 639.

[2] [2] H. Poincaré, in *Lorentz-Festschrift* (1900): 252-278.



velocity of light. K will maintain this velocity until the radiation complex, whose spatial extension is very small in comparison with the cavity of K, gets absorbed by B. The duration of the hollow cylinder's motion is (apart from terms of higher order) equal to a/V, if a denotes the distance from A to B. After absorption of the radiation complex by B, the body K is again at rest. During the radiation process under consideration, K has shifted a distance of

$$\delta = \frac{1}{V}\frac{S}{M} \cdot \frac{a}{V}$$

to the left.

In the cavity of K, let us have a body k (imagined as massless for the sake of simplicity) next to a (likewise massless) mechanism that can move the body k, which shall first be located in B, back and forth between B and A. After the amount of radiation S has been absorbed by B, this amount of energy shall be transferred to k, and then k moved to A. Finally, the amount of energy S shall again be taken up in A by the hollow cylinder K, and k shall be moved back to B again. The whole system has now undergone a complete cyclic process, which one can imagine to be repeated arbitrarily often.

If one assumes that the carrier body k remains massless even after it has absorbed the amount of energy S, then one also has to assume that the return transport of the amount of energy S is not associated with a change in position of the hollow cylinder K. Thus the only outcome of the entire cyclic process is a shift δ of the whole system to the left; by repeating the cyclic process, one can make this shift as large as desired. We thus arrive at the result that an initially stationary system can change the position of its center of gravity arbitrarily greatly without having external forces acting upon it, and without undergoing any permanent change.

It is clear that the result does not contain any inner contradictions; however, it does contradict the laws of mechanics, according to which a body originally at rest cannot perform a translational motion if no other bodies act upon it.

However, if one assumes that any energy E possesses the inertia E/V^2, then the contradiction with the principles of mechanics disappears. For according to this assumption the carrier body has a mass S/V^2 while it transports the energy amount S from B to A; and since the center of gravity of *the entire system* must be at rest during that process according to the center-of-mass theorem, the cylinder K undergoes during it a total shift S' to the right, amounting to

$$\delta' = a \cdot \frac{S}{V^2} \cdot \frac{1}{M} .$$

Comparison with the result found above shows that (at least in first approximation) $\delta = \delta'$, i.e., that the position of the system is the same before and after the cyclic process. This eliminates the contradiction with the principles of mechanics.

§2. *On the principle of the conservation of the motion of the center of gravity*

We consider a system of n discrete material points with masses $m_1, m_2 \ldots m_n$ and center of gravity coordinates $x_1 \ldots z_n$. With respect to thermal and electric phenomena, these material points are not to be conceived as elementary structures (atoms, molecules), but as bodies in the usual sense of small dimensions, whose energy is not determined by the velocity of the center of gravity. These masses could act on each other through electromagnetic processes as well as through conservative forces (i.e., gravity, rigid connections); however, we shall assume that both the potential energy of the conservative forces and the kinetic energy of the motion of the center of gravity of the masses are infinitesimally small relative to the "internal" energy of the masses $m_1 \ldots m_n$.

Assume that the Maxwell-Lorentz equations

$$(1)\quad \left\{\begin{aligned} \frac{u}{V}\rho + \frac{1}{V}\frac{dX}{dt} &= \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z}, \\ \frac{u}{V}\rho + \frac{1}{V}\frac{dY}{dt} &= \frac{\partial L}{\partial z} - \frac{\partial N}{\partial x}, \\ \frac{u}{V}\rho + \frac{1}{V}\frac{dZ}{dt} &= \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}, \\ \frac{1}{V}\frac{dL}{dt} &= \frac{\partial Y}{\partial z} - \frac{\partial Z}{\partial y}, \\ \frac{1}{V}\frac{dM}{dt} &= \frac{\partial Z}{\partial x} - \frac{\partial X}{\partial z}, \\ \frac{1}{V}\frac{dN}{dt} &= \frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x} \end{aligned}\right.$$ [3]

hold in the entire space, where

$$\rho = \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}$$

denotes the 4π-fold density of electricity.

If one adds up equations (1) after they have been successively multiplied by

$$\frac{V}{4\pi}Xx,\ \frac{V}{4\pi}Yx\ \ldots\ \frac{V}{4\pi}Nx$$

and integrates them over the entire space, one obtains, after a few integrations by parts, the following equation

$$(2)\quad \int \frac{\rho}{4\pi}\, x(uX + vY + wZ)\,d\tau + \frac{d}{dt}\left\{\int x \cdot \frac{1}{8\pi}(X^2 + Y^2\ \ldots + N^2)\,d\tau\right] - \frac{V}{8\pi}\int (YN - ZM)\,d\tau = 0.$$ [4]

The first term of this equation represents the energy supplied by the electromagnetic field to the bodies $m_1 \ldots m_n$. According to our hypothesis on the dependence of the masses on energy, the first term of the sum should therefore be equated with the expression [5]

$$V^2 \sum x_\nu \frac{dm_\nu}{dt},$$

since we assume according to the above that the individual material points m_ν change their energy, and thereby also their mass, *only* by taking up electromagnetic energy.

If we assign to the electromagnetic field too a mass density (ρ_e), which differs by a factor $1/V^2$ from the energy density, then the second term of the equation takes the form

$$V^2 \frac{d}{dt} \left\{ \int x\rho_e d\tau \right\} .$$

If the integral in the third term of equation (2) is denoted by J, then this equation becomes

$$\text{(2a)} \qquad \sum \left[x_\nu \frac{dm_\nu}{dt} \right] + \frac{d}{dt} \left\{ \int x\rho_e d\tau \right\} - \frac{1}{4\pi V} J = 0 .$$

We now have to find the meaning of the integral J. If one successively multiplies the second, third, fifth, and sixth of equations (1) by NV, $-MV$, $-ZV$, YV, adds them and integrates over the space, one obtains, after a few integrations by parts,

[6] $$\text{(3)} \qquad \frac{dJ}{dt} = - 4\pi V \int \frac{\rho}{4\pi} \left[X + \frac{v}{V} N - \frac{w}{V} M \right] d\tau = - 4VR_x ,$$

where R_x is the algebraic sum of the X-components of all forces exerted by the electromagnetic field upon the masses $m_1 \ldots m_n$. Since the corresponding sum of all forces due to the conservative interactions vanishes, R_x is at the same time the sum of the X-components of *all* forces acting upon the masses m_ν.

Next we shall consider equation (3), which is independent of the hypothesis that the mass depends on energy. If we disregard the dependence of the masses on energy and denote the resultant of all X-components of the forces acting on m_ν by $\mathfrak{X}_\nu$, we must set up the following equation of motion for the mass m_ν:

$$m_\nu \frac{d^2 x_\nu}{dt^2} = \frac{d}{dt}\left\{m_\nu \frac{dx_\nu}{dt}\right\} = \mathfrak{X}_\nu \ ; \tag{4}$$

hence we also obtain

$$\frac{d}{dt} \sum \left[m_\nu \frac{dx_\nu}{dt}\right] = \sum \mathfrak{X}_\nu = R_x \ . \tag{5}$$

From equation (5) and equation (3) one obtains

$$\frac{J}{4\pi V} + \sum m_\nu \frac{dx_\nu}{dt} = \text{const.} \tag{6}$$

If we reintroduce the hypothesis that the quantities m_ν depend on energy, and thus also on time, then we face the difficulty that the mechanical equations for that case are no longer known; the first equal sign of equation (4) thus does not hold anymore. However, one should take into consideration that the difference

$$\frac{d}{dt}\left\{m_\nu \frac{dx_\nu}{dt}\right\} - m_\nu \frac{d^2 x_\nu}{dt^2} = \frac{dm_\nu}{dt}\frac{dx_\nu}{dt} = \frac{1}{V^2}\int \frac{\rho}{4\pi}\frac{dx_\nu}{dt}(uX + vY + wZ)\,d\tau$$

is of second order in the velocities. Hence, if all velocities are so small that terms of second order may be neglected, then even if the mass m_ν is variable, the equation

$$\frac{d}{dt}\left[m_\nu \frac{dx_\nu}{dt}\right] = \mathfrak{X}_\nu$$

certainly holds with the required accuracy. Then equations (5) and (6) hold as well, and one obtains from equations (6) and (2a):

$$\frac{d}{dt}\left[\sum (m_\nu x_\nu) + \int x\rho_e\,d\tau\right] = \text{const.} \tag{2b}$$

If ξ denotes the X-coordinate of the center of gravity of the ponderable masses and of the energy mass of the electromagnetic field, then we have

$$\xi = \frac{\sum (m_\nu x_\nu) + \int x\rho_e d\tau}{\sum m_\nu + \int \rho_e d\tau},$$

where, according to the energy principle, the value of the denominator on the right-hand side is independent of time.[1] Hence we may write equations (2b) also in the form

$$\frac{d\xi}{dt} = \text{const.} \tag{2c}$$

Thus, if one ascribes the inertial mass E/V^2 to any energy E, then—at least in first approximation—the principle of conservation of the motion of the center of gravity also holds for systems in which electromagnetic processes take place.

The present investigation shows that one either has to give up the fundamental law of mechanics, according to which a body originally at rest cannot perform translational motion unless acted upon by external forces, or one has to assume that a body's inertia depends on its energy content according to the law stated.

Bern, May 1906. (Received on 17 May 1906)

---

[1] According to the interpretation developed in this paper, the principle of the constancy of mass is a special case of the energy principle.

Doc. 36

ON A METHOD FOR THE DETERMINATION OF THE RATIO OF THE TRANSVERSE AND THE LONGITUDINAL MASS OF THE ELECTRON

by A. Einstein

[*Annalen der Physik* 21 (1906): 583-586]

Three quantities concerning cathode rays are accessible to precise observation: the potential difference producing the velocity of the rays (generating potential), the electrostatic deflection, and the magnetic [1] deflection. There exist two independent relations between these three quantities whose knowledge at considerable ray velocities is of extraordinary theoretical interest. One of these relations, namely that between magnetic and electrostatic deflection, has been examined for β-rays by Mr. Kaufmann. [2]

In the following I shall point out that there exists one other relation between these quantities that can be measured with sufficient accuracy, namely, that between the generating potential and the electrostatic deflection of cathode rays, or, what is the same, the ratio of the transverse to the longitudinal electron mass as a function of the generating potential. [3]

If the square of the velocity of the electrons is very small compared with the square of the velocity of light, the motion of the electron obeys the equations

$$\frac{d^2x}{dt^2} = -\frac{\epsilon}{\mu_0} X, \quad \text{etc.},$$

where ϵ/μ_0 denotes the ratio of the charge to the mass of the electron, x, y, z the coordinates of the electron, and X, Y, Z the components of the electric field strength if no other forces besides the electrostatic ones act on the electron. We assume that the electrons move with an initial velocity zero from some starting point x_0, y_0, z_0 (cathode). The motion is then uniquely determined by the equations given above; it shall be given by the equations

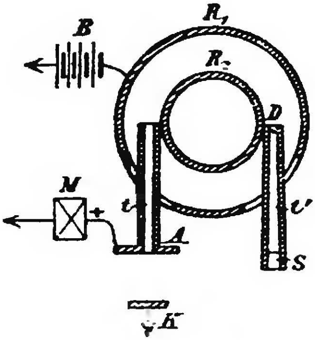
$$x = \varphi_1(t) \ ,$$
$$y = \varphi_2(t) \ ,$$
$$z = \varphi_3(t) \ .$$

If one imagines all electrostatic force components to be multiplied everywhere by n^2, then, as can be seen easily from the above equations of motion, the electron will move according to the equations

$$x = \varphi_1(nt) \ ,$$
$$y = \varphi_2(nt) \ ,$$
$$z = \varphi_3(nt) \ .$$

From this it follows that a proportional change of the field is accompanied by a change of the electrons' velocity but not of their trajectory.

A change of trajectory evidently is produced by a proportional change of the field only at electron velocities at which the ratio of transverse to longitudinal mass is noticeably different from unity. If the electrostatic field is chosen such that the cathode rays travel a strongly curved path, then even small differences between the transverse and the longitudinal mass will have an observable effect on the trajectory. The accompanying sketch shows an arrangement by which the ratio of the transverse to the longtudinal mass of the electron could be determined on the basis of the principle indicated. The cathode rays attain their velocity between the grounded cathode K and the



anode A, which is attached to the positive terminal of the current source M and which serves at the same time as a shutter, and are then introduced, via a thin tube t connected with A, into the space between the metal cylinders R_1 and R_2. R_1 is grounded, R_2 is conductively connected with t, i.e., with the positive pole of the current source, whose negative pole is grounded.

The dimensions should be chosen such that slow cathode rays move approximately in a circle, at a short distance from R_2. The rays then enter the somewhat conical metal tube t', which is connected by metal with R_2 and inside which there is a phosphorescent screen S on which there shall fall the shadow of a vertical wire D set up at the interior end of t'.

When slow cathode rays are applied, the shadow of D on S takes up a quite definite position (zero position). If the rays' generating potential is increased, the shadow of the wire will shift. By inserting a battery B into the ground connection of R_1 the shadow shall be returned to the zero position.

If Π denotes the potential at which the shadow-forming rays get deflected, then Π is also the potential difference that imparts the kinetic energy to the deflected rays. Further, if ρ denotes the radius of curvature of the shadow-forming rays, then we have

$$\frac{\mu_t}{\mu_\ell} = \frac{\rho}{2}\frac{X}{\Pi} . \qquad [4]$$

Here μ_t denotes the "transverse mass" of the electron, μ_ℓ that longitudinal mass which is defined by the equation

$$\text{kinetic energy} = \mu_\ell \frac{v^2}{2} ,$$

and X the deflecting electric force.

If P denotes the potential of R_2 (potential of the positive pole of the current source M), and p the potential of R_1 at which the shadow is in the zero position, then

$$\Pi = P - \alpha(P - p) ,$$

where α denotes a constant that depends on the dimensions of the apparatus and is small compared with 1. Further, the quantity X is proportional to [5] the potential difference $P - p$. Thus, one obtains from the above equation

$$\frac{\mu_t}{\mu_\ell} = \text{const.}\ \frac{P - p}{P - \alpha(P - p)}\ ,$$

or (with some permissible approximations)

$$\frac{\mu_t}{\mu_\ell} = \text{const.}\left[1 - (1 + \alpha)\frac{p}{P}\right]\ .$$

Since α can evidently be obtained with sufficient accuracy, and P and p can be measured accurately within a few percent, the accuracy with which one can obtain the deviation of μ_t/μ_ℓ from unity is basically determined by the accuracy with which one can set the zero position of the wire shadow. One can easily see that the latter accuracy can be made sufficiently great that a 0.3% deviation of the quantity μ_t/μ_ℓ from unity (which corresponds to a shift of the shadow of about 1 mm when $\overline{DS} = 10$ cm) can still be noticed. It is especially worth mentioning that the accuracy of measurement is not significantly affected by the unavoidable fluctuations of the potential P during the experiment. [6]

Finally we would like to give the relation between μ_t/μ_ℓ and Π in first approximation, as obtained from different theories. If Π is expressed in volts, then we have

[7] according to the theory of Bucherer:

$$\frac{\mu_t}{\mu_\ell} = 1 - 0.0070 \cdot \frac{\Pi}{10,000}\ ,$$

[8] according to the theory of Abraham:

$$\frac{\mu_t}{\mu_\ell} = 1 - 0.0084 \cdot \frac{\Pi}{10,000}\ ,$$

[9] according to the theory of Lorentz and Einstein:

$$\frac{\mu_t}{\mu_\ell} = 1 - 0.0104 \cdot \frac{\Pi}{10,000}\ ,$$

Since I am not in a position to do experimental work myself, I would be glad if a physicist would show an interest in the method described. [10]

Bern, August 1906. (Received on 4 August 1906)

Doc. 37

Review of M. PLANCK, "*Lectures on the theory of thermal radiation*" [1]
("*Vorlesungen über die Theorie der Wärmestrahlung*," Leipzig:
J. A. Barth, 1906. 222 pp. 7.80 mark)
[*Beiblätter zu den Annalen der Physik* 30 (1906): 764-766]

In the book under consideration the fundamental works of Kirchhoff, W. Wien and the author have been united into a whole of marvelous clarity and [2] unity, so that the book is superbly suited for familiarizing the reader fully with the material—even if the area dealt with has been totally unfamiliar to him.

In the first section (pp. 1-23) the basic concepts and terms (such as "emission coefficient," "coefficient of diffusion," "reflecting surface," "smooth" and "rough surface," "black surface," "black body," "coefficient of absorption," "pencil of rays," "intensity," "radiation density," etc.) are first defined and—insofar as they are definitionally interrelated—linked together mathematically. Then (pp. 23-48) the Clausius relation concerning [3] the ratio of radiation densities in media with different indices of refraction as well as the Kirchhoff relation between emissivity and absorptivity are derived. [4]

While up to this point only the laws of ray optics have been employed, the second section (pp. 49-99) employs the Maxwell theory, though exclusively for the derivation of the radiation pressure. The magnitude of the latter, as the author emphasizes, cannot be obtained from considerations based on energetics. With the aid of the expression obtained for the radiation pressure, the [5] Stefan-Boltzmann law and the Wien displacement law are derived, and the concepts "temperature of monochromatic radiation" and "temperature of a monochromatic elementary pencil of rays" are defined.

The Wien displacement law yields for the energy density u in the normal spectrum the equation $u = \nu^3\varphi(T/\nu)$, where T denotes the absolute temperature and ν the frequency. Sections three and four of the book (pp. 100-179) contain an exposition of the author's fundamental investigations aimed at the determination of the function φ that appears in the Wien displacement law. Even though this goal could not be achieved in a purely

[6] deductive way, only using theoretical aids adequately supported empirically [7] —the author uses a hypothesis supported only by analogy—every impartial reader will find that a high probability attaches to the result obtained.

The course of the investigation is as follows: First, the oscillation equation of a resonator of small dimensions and small damping, located in a radiation field, is established on the basis of Maxwell's equations. Then one determines the mean energy of a resonator in a stationary radiation field with the aid of the oscillation equation, and, using the second law, the "temperature of the resonator" as function of the above universal function. [8] This reduces the problem of energy distribution in the normal spectrum to the task of determining the entropy of a system consisting of a large number of radiation resonators of the same frequency.

To solve the latter problem, it is first explained, based on Boltzmann's [9] works, that one is led to a correct determination of the entropy S if one puts $S = k \log W$, where k denotes a (universal) constant and W the number of "complexions." The latter quantity represents the multiplicity of all those possible distributions of the elementary variables that belong to the complex of observed quantities to which the entropy S corresponds.

In order to be able to determine the quantity W by counting, one must divide the whole available region of the state variables into discrete elementary regions. In general, the result depends on the absolute magnitudes as [10] well as on the ratios of the magnitudes of these elementary regions. While for the determination of the quantity W of a resonator system one chooses the magnitude *ratio* of the elementary regions as in a sinusoidally oscillating structure in the theory of gases, one chooses—in contrast to the assumption on infinitesimally small elementary regions generally used until now in the [11] theory of gases—the elementary regions to be of finite magnitude ($= h\nu$), where ν denotes the frequency and h a universal constant; $h\nu$ has the dimension of energy. The author points repeatedly to the necessity of introducing this universal constant h and emphasizes the importance of a physical interpretation (not given in the book) of the latter.

From the expression for the entropy S, obtained in the way indicated, one then derives the familiar Planck radiation formula,

$$u = \frac{8\pi h\nu^3}{c^3} \cdot \frac{1}{e^{h\nu/k\cdot T} - 1} .$$

The fourth section contains, further, Planck's determination of the elementary quanta, as well as discussions of works of various authors on radiation theory. [12]

The last section of the book (pp. 180-222), which deals with irreversible radiation processes, offers deep insight into the nature of the irreversibility of thermal processes.

Doc. 38

PLANCK'S THEORY OF RADIATION AND THE THEORY OF SPECIFIC HEAT

by A. Einstein

[*Annalen der Physik* 22(1907): 180-190]

In two previous papers[1] I have shown that the interpretation of the law of energy distribution of black-body radiation in terms of Boltzmann's theory of the second law leads to a new conception of the phenomena of light emission and light absorption, which, even though still far from having the character of a complete theory, is remarkable insofar as it facilitates the understanding of a series of regularities. The present paper will show that the theory of radiation—in particular Planck's theory—leads to a modification of the molecular-kinetic theory of heat by which some difficulties obstructing the implementation of that theory can be eliminated. The paper will also yield a relationship between the thermal and optical behavior of solids. [2]

First we will give a derivation of the mean energy of Planck's resonator that clearly demonstrates its relation to molecular mechanics.

To that end we use a few results of the general molecular theory of heat.[1] Let the state of a system in the sense of the molecular theory be [3] completely determined by the (very many) variables $P_1, P_2 \ldots P_n$. Let the molecular process proceed according to the equations

$$\frac{dP_\nu}{dt} = \Phi_\nu(P_1, P_2 \; . \; . \; . \; P_n), \qquad (\nu = 1, 2 \; . \; . \; . \; n) \; ,$$

and let the relation

[4] (1)
$$\sum \frac{\partial \Phi_\nu}{\partial P_\nu} = 0$$

hold for all values of the P_ν's.

---

[1] [1] A. Einstein, *Ann. d. Phys.* 17 (1905): 132 and 20 (1905): 199.

Further, let a partial system of the system of the P_ν's be determined by the variables $p_1 \ldots p_n$ (which belong to the P_ν's), and let it be assumed that the energy of the whole system can with good approximation be thought of as composed of two parts, of which one (E) depends on the $p_1 \ldots p_m$ *only*, while the other is independent of $p_1 \ldots p_m$. Also, let E be infinitesimally small compared with the total energy of the system.

The probability dW that at a randomly picked instant the p_ν's lie in an infinitesimally small region $(dp_1, dp_2 \ldots dp_m)$ is then given by the equation[1]

$$dW = Ce^{-\frac{N}{RT}E} dp_1 \ldots dp_m \,. \tag{2}$$ [6]

Here C is a function of the absolute temperature (T), N is the number of molecules in one gram-equivalent, R is the constant of the gas equation referring to one gram-molecule.

If one puts

$$\int_{dE} dp_1 \ldots dp_m = \omega(E)dE \,,$$

where the integral is to be extended over all combinations of the P_ν's to which correspond energy values between E and $E + dE$, one obtains

$$dW = Ce^{-\frac{N}{RT}E} \omega(E)dE \,. \tag{3}$$

If one chooses as the variables P_ν the center-of-mass coordinates and velocity components of mass points (atoms, electrons) and assumes that the accelerations depend only on the coordinates, but not on the velocities, then one arrives at the molecular-kinetic theory of heat. The relation (1) is here satisfied, so that equation (2) holds as well.

In particular, if one imagines that one has chosen as the system of the p_ν's, an elementary mass particle which can perform sinusoidal oscillations

---

[1] A. Einstein, *Ann. d. Phys.* 11 (1903): 170ff. [5]

along a straight line, and denotes its instantaneous distance from the equilibrium position and velocity by x and ξ, respectively, one obtains

$$dW = Ce^{-\frac{N}{RT}E}\,dx\,d\xi \,, \tag{2a}$$

and since one has to take $\int dx\,d\xi = \text{const.}\,dE$, hence $\omega = \text{const.}$[1]:

$$dW = \text{const.}\; e^{-\frac{N}{RT}E}\,dE \,. \tag{3a}$$

The mean value of the mass particle's energy is therefore

$$\bar{E} = \frac{\int Ee^{-\frac{N}{RT}E}\,dE}{\int e^{-\frac{N}{RT}E}\,dE} = \frac{RT}{N} \,. \tag{4}$$

It is obvious that formula (4) can also be applied to a rectilinearly oscillating ion. If one does so, and takes into account that, according to a study by Planck[2], the relation

[8] $$\bar{E}_\nu = \frac{L^3}{8\pi\nu^2}\,\rho_\nu \tag{5}$$

must hold between its mean energy $\bar{E}$ and the density ρ_ν of the black-body radiation at the frequency considered there, then by eliminating $\bar{E}$ from (4) and (5) one arrives at Rayleigh's formula

[9] $$\rho_\nu = \frac{R}{N}\,\frac{8\pi\nu^2}{L^3}\,T \,, \tag{6}$$

which, as is well known, represents only a limiting law for large values of [10] T/ν.

---

[1] Because one has to set $E = ax^2 + b\xi^2$.

[7] [2] M. Planck, *Ann. d. Phys.* 1 (1900): 99.

To arrive at Planck's theory of black-body radiation, one can proceed as follows.[1] One retains equation (5), i.e., one assumes that Maxwell's theory [12] of electricity yields the correct relationship between radiation density and $\bar{E}$. On the other hand, one abandons equation (4), i.e., one assumes that it is the application of the molecular-kinetic theory which causes a conflict with experience. However, we maintain the formulas (2) and (3) of the general molecular theory of heat. Instead of setting

$$\omega = \text{const.}$$

in accordance with the molecular-kinetic theory, we set $\omega = 0$ for all values of E that are not extremely close to 0, ϵ, 2ϵ, 3ϵ, etc. Only between 0 and $0 + \alpha$, ϵ and $\epsilon + \alpha$, 2ϵ and $2\epsilon + \alpha$, etc. (where α is infinitesimally small compared with ϵ) shall ω be different from zero, such that

$$\int_0^{\alpha} \omega dE = \int_{\epsilon}^{\epsilon+\alpha} \omega dE = \int_{2\epsilon}^{2\epsilon+\alpha} \omega dE = \ldots = A$$

As can be seen from equation (3), this stipulation involves the assumption that the energy of the elementary structure under consideration assumes only values that are infinitesimally close to 0, ϵ, 2ϵ, etc.

Using the above stipulation for ω, one obtains with the help of (3):

$$\bar{E} = \frac{\int Ee^{-\frac{N}{RT}E}\omega(E)dE}{\int e^{-\frac{N}{RT}E}\omega(E)dE} = \frac{0 + A\epsilon e^{-\frac{N}{RT}\epsilon} + A.2\epsilon e^{-\frac{N}{RT}2\epsilon} \ldots}{A + Ae^{-\frac{N}{RT}\epsilon} + Ae^{-\frac{N}{RT}2\epsilon} + \ldots} = \frac{\epsilon}{e^{-\frac{N}{RT}\epsilon} - 1} .$$ [13]

If one also sets $\epsilon = (R/N)\beta\nu$ (according to the quantum hypothesis), one [14] obtains from this

---

[1] Cf. M. Planck, *Vorlesungen über die Theorie der Wärmestrahlung* [Lectures on the theory of thermal radiation]. (Leipzig: J.A. Barth, 1906), §§149, 150, 154, 160, 166. [11]

[15] (7)
$$\bar{E} = \frac{\frac{R}{N}\beta\nu}{e^{\frac{\beta\nu}{T}} - 1}$$

as well as, with the help of (5), the Planck radiation formula,

$$\rho_\nu = \frac{8\pi}{L^3} \cdot \frac{R\beta}{N} \frac{\nu^3}{e^{\frac{\beta\nu}{T}} - 1} .$$

Equation (7) shows the dependence of the mean energy of Planck's resonator on the temperature.

From the above it emerges clearly in which sense the molecular-kinetic theory of heat must be modified in order to be brought into agreement with the distribution law of black-body radiation. For although one has thought before that the motion of molecules obeys the same laws that hold for the motion of bodies in our world of sense perception (in essence, we are only adding the postulate of complete reversibility), we now must assume, for ions capable of oscillating at particular frequencies which can mediate an exchange of energy between matter and radiation, that the diversity of states they can assume is
[16] less than for bodies within our experience. For we had to make the assumption that the mechanism of energy transfer is such that the energy of elementary structures can only assume the values 0, $(R/N)\beta\nu$, $2(R/N)\beta\nu$, etc.[1]

I believe that we must not content ourselves with this result. For the question arises: If the elementary structures that are to be assumed in the theory of energy exchange between radiation and matter cannot be perceived in terms of the current molecular-kinetic theory, are we then not obliged also to modify the theory for the other periodically oscillating structures considered in the molecular theory of heat? In my opinion the answer is not in doubt. If Planck's radiation theory goes to the root of the matter, then contradic-

---

[1] It is obvious that this assumption also has to be extended to bodies capable of oscillation that consist of any number of elementary structures.

tions between the current molecular-kinetic theory and experience must be expected in other areas of the theory of heat as well, which can be resolved along the lines indicated. In my opinion this is actually the case, as I shall now attempt to show.

The simplest conception one can form about thermal motion in solids is that its individual atoms perform sinusoidal oscillations about equilibrium positions. With this assumption, by applying the molecular-kinetic theory [17] (equation (4)) while taking into account that three degrees of freedom of motion must be assigned to each atom, one obtains for the specific heat of a gram-equivalent of the substance

$$c = 3Rn \ ,$$

or—expressed in gram-calories—

$$c = 5.94\ n \ ,$$

when n denotes the number of atoms in the molecule. It is well known that this relation applies with remarkably close approximation to most elements and to many compounds in the solid aggregation state (Dulong-Petit's law, rule of F. Neumann and Kopp). [18]

However, if one examines these facts a little closer, one encounters two difficulties that seem to set narrow limits on the applicability of the molecular theory.

1. There are elements (carbon, boron, and silicon) that in the solid state and at ordinary temperatures have specific atomic heats much smaller than 5.94. Furthermore, the specific heat per gram-molecule is less than [19] $n \cdot 5.94$ in all solid compounds containing oxygen, hydrogen or at least one of the elements just mentioned. [20]

2. Mr. Drude has shown[1] that the optical phenomena (dispersion) lead to the conclusion that several elementary masses moving independently of each other must be ascribed to each atom of a compound in that he successfully

---

[1] P. Drude, *Ann. d. Phys.* 14 (1904): 677. [21]

related the infrared proper frequencies to oscillations of atoms (atom ions) [22] and the ultraviolet proper frequencies to the oscillations of electrons. This poses a second significant difficulty for the molecular-kinetic theory of heat, because the specific heat would have to exceed significantly the value 5.94 n, since the number of mobile mass points per molecule is larger than the latter's number of atoms.

Based on the above one should note here the following: If we conceive of the carriers of heat in solids as periodically oscillating structures whose frequency is independent of their oscillation energy, then according to Planck's theory of radiation we should not expect the value of the specific heat always to be 5.94 n. Rather, we have to set (7)

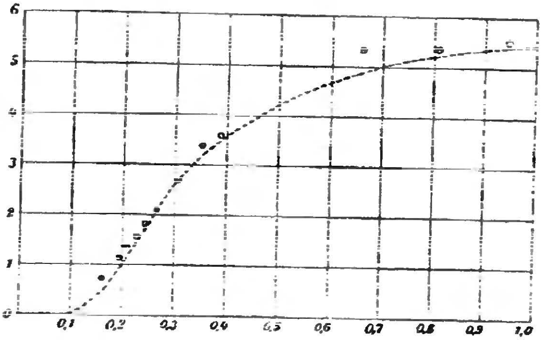
$$\bar{E} = \frac{3R}{N} \frac{\beta\nu}{e^{\frac{\beta\nu}{T}} - 1} .$$

The energy of N such elementary structures, measured in gram-calories, hence has the value

$$5.94 \frac{\beta\nu}{e^{\frac{\beta\nu}{T}} - 1} ,$$

so that each such oscillating elementary structure contributes to the specific heat the value

$$5.94 \frac{e^{\frac{\beta\nu}{T}} \left(\frac{\beta\nu}{T}\right)^2}{\left(e^{\frac{\beta\nu}{T}} - 1\right)^2} \tag{8}$$



per gram-equivalent. Thus, summation over all species of oscillating elementary structures occurring in the solid substance in question yields the following expression for the specific heat per gram-equivalent[1]:

$$c = 5.94 \sum \frac{e^{\frac{\beta\nu}{T}} \cdot \left[\frac{\beta\nu}{T}\right]^2}{\left[e^{\frac{\beta\nu}{T}} - 1\right]^2} . \tag{8a}$$

The above figure[2] shows the value of expression (8) as a function of $x = (T/\beta\nu)$. If $(T/\beta\nu) > 0.9$, the contribution of the structure to the specific molecular heat does not differ significantly from the value 5.94, which also follows from the heretofore accepted molecular-kinetic theory; the smaller the ν, the lower the temperature at which this will already be the case. In contrast, if $(T/\beta\nu) < 0.1$, the elementary structure in question does not contribute significantly to the specific heat. In between, the expression (8) initially grows faster and then more slowly.

From what has been said it follows first of all that the electrons capable of oscillation, which have to be postulated to explain the ultraviolet proper frequencies, cannot significantly contribute to the specific heat at normal temperatures $(T = 300)$, because the inequality $(T/\beta v) < 0.1$ becomes [23] the inequality $\lambda < 4.8\ \mu$ at $T = 300$. On the other hand, if the elementary structure satisfies the condition $\lambda > 48\mu$, then according to what has been said above, its contribution to the specific heat must be close to 5.94 at usual temperatures.

Since generally for infrared proper frequencies $\lambda > 4.8\mu$, according to [24] our conceptions these proper oscillations must contribute to the specific heat, and the greater the λ, the greater this contribution. According to Drude's investigations, these proper frequencies are to be attributed to the [25] ponderable atoms (atom ions) themselves. The most obvious conclusion seems therefore to be to consider exclusively the positive atom ions as the carriers of heat in solids (insulators).

---

[1] This consideration can easily be extended to anisotropic bodies.

[2] Cf. dashed curve.

If the infrared proper oscillation frequencies ν of a solid are known, then according to the aforesaid its specific heat as well as its dependence on [26] the temperature would be completely determined by equation (8a). Pronounced deviations from the relation $c = 5.94\ n$ would have to be expected at normal temperatures if the substance in question showed an optical infrared proper frequency for which $\lambda < 48\ \mu$; at sufficiently low temperatures the specific heats of all solid bodies should decrease significantly with decreasing [27] temperature. Further, the Dulong-Petit law as well as the more general law $c = 5.94\ n$ must hold for all bodies at sufficiently high temperatures unless new degrees of freedom of motion (electron-ions) become apparent at the [28] latter.

Both above-mentioned difficulties are resolved by the new interpretation and I believe it likely that the latter will prove its validity in principle. Of course, an exact agreement with the facts is out of the question. During [29] heating, solids experience changes in molecular arrangement (e.g., changes in volume) that are associated with changes in energy content; all solids that conduct electricity contain freely moving elementary masses that make a contribution to the specific heat; the random heat oscillations have possibly a somewhat different frequency than the proper oscillations of the elementary structures during optical processes. Finally, the assumption that the pertinent elementary structures have an oscillation frequency that is independent of the energy (temperature) is undoubtedly inadmissible.

Nevertheless, it is interesting to compare our conclusions with observation. Since we are concerned with rough approximations only, we assume, in accordance with F. Neumann-Kopp's rule, that every element contributes equally to the molecular specific heat of all its solid compounds even if its specific heat is abnormally small. The numerical data presented in the [30] following table are taken from Roskoe's textbook of chemistry. We note that all elements with abnormally low atomic heat have low atomic weights; according to our interpretation, this is to be expected, since, ceteris paribus, low atomic weights correspond to high oscillation frequencies. The last column of the table lists the values of λ in microns that are obtained from these numbers, if one assumes that they are valid at $T = 300$, with the help of the curve showing the relation between x and c.

Element	Specific atomic heat	$\lambda_{\text{calc.}}$
S and P	5.4	42
Fl	5	33
O	4	21
Si	3.8	20
B	2.7	15
H	2.3	13
C	1.8	12

Further, we take some data on infrared proper oscillations (metallic reflection, residual rays) of some transparent solids from the tables of Landolt and Börnstein; the observed λ are listed in the table below as "$\lambda_{\text{obs.}}$"; the numbers under "$\lambda_{\text{calc.}}$" are taken from the above table if they [31] refer to atoms with abnormally low specific heat; for the others it is assumed that $\lambda > 48\ \mu$.

Substance	$\lambda_{\text{obs.}}$	$\lambda_{\text{calc.}}$
CaFl	24; 31.6	33; >48
NaCl	51.2	>48
KCl	61.2	>48
$CaCO_3$	6.7; 11.4; 29.4	12; 21; >48
SiO_2	8.5; 9.0; 20.7	20; 21

In the table, NaCl and KCl contain only atoms with normal specific heat; indeed, the wavelengths of their infrared proper oscillations are larger than 48 μ. The other substances contain only atoms with abnormally low specific heats (except for Ca); indeed, the frequencies of these substances range between 4.8 and 48 μ. In general, the values of λ obtained theoretically from specific heats are considerably larger than those observed. It is possible that these deviations might be explained by a strong variation of the frequency of the elementary structure with its energy. Be that as it may, the agreement of the observed and the calculated λ is remarkable both with respect to the sequence as well as with respect to the order of magnitude.

Finally, we will also apply the theory to diamond. Its infrared proper frequency is not known, but can be calculated on the basis of the theory described if the molecular specific heat c is known for some temperature T;

the x corresponding to c can be taken directly from the curve, and λ is then calculated from the relation $(TL/\beta\lambda) = x$.

I am using the experimental results of H. F. Weber, which I took from [32] the tables of Landolt and Börnstein (cf. the following table). For $T = 331.3$ we have $c = 1.838$; according to the theory described, from this it follows that $\lambda = 11.0\ \mu$. Based on this value, those in the table's third column are calculated according to the formula $x = (TL/\beta\lambda)$, $(\beta = 4.86 \cdot 10^{-11})$.

T	c	x
222.4	0.762	0.1679
262.4	1.146	0.1980
283.7	1.354	0.2141
306.4	1.582	0.2312
331.3	1.838	0.2500
358.5	2.118	0.2705
413.0	2.661	0.3117
479.2	3.280	0.3615
520.0	3.631	0.3924
[33] 879.7	5.290	0.6638
1079.7	5.387	0.8147
1258.0	5.507	0.9493

The points, whose abscissas are these values of x and whose ordinates are the values of c as obtained experimentally from Weber's observations and listed in the table, should lie on the x,c-curve shown above. We plotted these points—indicated by circles—in the above figure; in fact, they do almost lie on the curve. Hence we have to assume that the elementary carriers of heat in diamond are almost monochromatic structures.

Thus, according to the theory it is to be expected that diamond shows an [34] absorption maximum at $\lambda = 11\ \mu$.

Bern, November 1906. (Received on 9 November 1906)

Doc. 39

ON THE LIMIT OF VALIDITY OF THE LAW OF THERMODYNAMIC EQUILIBRIUM AND ON THE POSSIBILITY OF A NEW DETERMINATION OF THE ELEMENTARY QUANTA

by A. Einstein

[*Annalen der Physik* 22 (1907): 569-572] [1]

Let the state of a physical system be determined in the thermodynamic sense by parameters λ, μ, etc. (e.g., readings of a thermometer, length or volume of a body, amount of a substance of a certain kind in one phase). If, as we assume, the system is not interacting with other systems, then, according to the laws of thermodynamics, equilibrium will occur at particular values λ_0, μ_0, etc. of the parameters, for which the system's entropy S is a maximum. However, according to the molecular theory of heat, this is not exactly but only approximately correct; according to this theory, the value of the parameter λ is not constant even at temperature equilibrium, but shows irregular fluctuations, though it is very rarely much different from λ_0.

At first glance the theoretical examination of the statistical law that governs these fluctuations would seem to require that certain stipulations regarding the molecular model must be applied. However, this is not the case. [2] Rather, essentially it is sufficient to apply the well-known Boltzmann relation connecting the entropy S with the statistical probability of a state. As we know, this relation is

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

where R is the constant of the gas equation and N is the number of molecules in one gram-equivalent.

We consider a state of the system in which the parameter λ has a value $\lambda_0 + \epsilon$ differing very little from λ_0. To bring the parameter λ from the value λ_0 to the value λ along a reversible path at constant energy E, one will have to supply some work A to the system and to withdraw the corresponding amount of heat. According to thermodynamic relations, we have

$$A = \int dE - \int TdS ,$$

or, since the change in question is infinitesimally small and $\int dE = 0$,

$$A = -T(S - S_0) \ .$$

On the other hand, however, according to the connection between entropy and probability of state, we have

$$S - S_0 = \frac{R}{N} \lg \left[\frac{W}{W_0}\right] \ .$$

From the last two equations it follows that

$$A = -\frac{RT}{N} \lg \frac{W}{W_0}$$

or

$$W = W_0 e^{-\frac{N}{RT} A} \ .$$

The result involves a certain degree of inaccuracy, because in fact one cannot talk about the probability of a *state*, but only about the probability [3] of a state *range*. If instead of the equation found we write

$$dW = \text{const.}\ e^{-\frac{N}{RT} A}\, d\lambda \ ,$$

then the latter law is exact. The arbitrariness due to our having inserted the differential of λ rather than the differential of some function of λ [4] into the equation will not affect our result.

We now put $\lambda = \lambda_0 + \epsilon$ and restrict ourselves to the case that A can be developed in positive powers of ϵ, and that only the first nonvanishing term of this series contributes noticeably to the value of the exponent at such values of ϵ for which the exponential function is still noticeably different from zero. Thus, we put $A = a\epsilon^2$ and obtain

$$dW = \text{const.}\ e^{-\frac{N}{RT} a\epsilon^2}\, d\epsilon \ .$$

Thus, in this case there applies the law of chance errors to the deviations ϵ. For the mean value of the work A one obtains

$$\bar{A} = \frac{1}{2} \frac{R}{N} T .$$

Hence, the mean value of the square of the fluctuation ϵ of a parameter λ is such that, in order to change the parameter λ from λ_0 to $\lambda^0 + \sqrt{\overline{\epsilon^2}}$ at constant energy of the system, the external work A that one would have to apply, if thermodynamics were strictly valid, equals $\frac{1}{2} \frac{R}{N} T$ (i.e., one-third of the mean kinetic energy of one atom).

If one inserts the numerical values for R and N, one obtains approximately

$$\bar{A} = 10^{-16} T .$$

We shall now apply the result obtained to a short-circuited condenser of (electrostatically measured) capacitance c. If $\sqrt{\overline{p^2}}$ is the mean (electrostatic) potential difference that the condenser assumes as a result of molecular disorder, then

$$\bar{A} = \frac{1}{2} c\overline{p^2} = 10^{-16} T .$$

We assume that the condenser is an air condenser consisting of two interlocking plate systems containing 30 plates each. The average distance between each plate and the adjacent plate of the other system shall be 1 mm. The size of the plates shall be 100 cm². The capacitance c is then about 5,000. At normal temperature one then obtains [5]

$$\sqrt{\overline{p^2}_{\text{stat.}}} = 3.4 \times 10^{-9} .$$ [6]

Measured in volts, one obtains

[7]
$$\sqrt{\overline{p^2_{\text{volt}}}} = 10^{-6} \ .$$

If one imagines that the two plate systems can move relative to one other, so that they can be completely separated, one can get the capacitance to be of order of magnitude 10 after the plates have been moved apart. If π denotes the potential difference resulting from p due to the separation, one obtains

$$\sqrt{\overline{\pi^2}} = 10^{-6} \cdot \frac{5{,}000}{10} = 0.0005 \text{ volt} \ .$$

Thus, if the condenser is short-circuited when the plate systems are pushed together, and the plates are pulled apart after the connection has been broken, potential differences of the order of magnitude of one-half millivolt will result between the plate systems.

It does not seem to me out of the question that these potential differ-
[8] ences may be accessible to measurement. For if metal parts can be electrically connected and separated without the occurrence of other *irregular* potential differences of the same order of magnitude as those calculated above, then it must be possible to achieve the goal by combining the above plate condenser
[9] with a multiplier. We would then have a phenomenon akin to Brownian motion in the domain of electricity that could be used for the determination of the quantity N.

Bern, December 1906. (Received on 12 December 1906)

Doc. 40

THEORETICAL REMARKS ON BROWNIAN MOTION

by A. Einstein

[*Zeitschrift für Elektrochemie und angewandte physikalische Chemie* 13 (1907): 41-48]

Prompted by Svedberg's investigation on the motion of suspended particles, published recently in the *Z. f. Elektroch.*, I consider it appropriate to [1] call attention to some properties of this motion required by the molecular theory. I hope that the following remarks will make it somewhat easier for the physicists who study this problem experimentally to interpret their observational data and to compare them with theory. [2]

1. The molecular theory of heat allows the calculation of the mean value of the instantaneous velocity a particle possesses at the absolute temperature T, since the kinetic energy of the particle's center-of-gravity motion is independent of the size and nature of the particle and of the nature of its environment, e.g., of the liquid in which the particle is suspended; this kinetic energy is equal to that of a monoatomic gas molecule. The mean velocity $\sqrt{\overline{v^2}}$ of a particle of mass m is threfore determined by the equation

$$m\frac{\overline{v^2}}{2} = \frac{3}{2}\frac{RT}{N},$$ [3]

where $R = 8.3\cdot10^7$, T is the absolute temperature, and N is the number of real molecules in a gram-molecule (about $4\cdot10^{23}$). We shall calculate $\sqrt{\overline{v^2}}$, as [4] well as the other quantities to be considered below, for particles of colloidal solutions of platinum studied by Mr. Svedberg. For these particles [5] we have to put $m = 2.5\cdot10^{-15}$, so that we get for $T = 292$

$$\sqrt{\overline{v^2}} = \sqrt{\frac{3RT}{mN}} = 8.6 \text{ cm/sec.}$$ [6]

2. We will now examine whether there is any chance of actually observing this enormous velocity on a suspended particle.

If we did not know anything about the molecular theory of heat, we would expect the following: If we imparted a velocity to a particle suspended in a liquid by an impulse of an external force, this velocity would be rapidly used up through the friction of the liquid. We neglect the latter's inertia and bear in mind that the resistance experienced by the particle moving with velocity v is $6\pi kPv$, where k denotes the coefficient of viscosity of the liquid and P the radius of the particle. We get the equation

[7] $$m \frac{dv}{dt} = -6\pi kPv.$$

This yields for the time ϑ in which the velocity decreases to one tenth of its initial value

$$\vartheta = \frac{m}{0.434 \cdot 6\pi kP} .$$

For the platinum particle (in water) mentioned above, we have to put
[8] $P = 2.5 \cdot 10^{-6}$cm, and $\eta = 0.01$, so that we get[1]

$$\vartheta = 3.3 \cdot 10^{-7} \text{ seconds}.$$

Returning to the molecular theory of heat, we must modify this analysis. True, we must assume now as well that, due to friction, the particle loses almost all its initial motion during the very short time ϑ. But we also must assume that during this time the particle receives new impulses by a process that is the reverse of internal friction, so that it retains a velocity that on the average equals $\sqrt{\overline{v^2}}$. But since we must imagine that the direction and magnitude of these impulses are (almost) independent of the initial direction of motion and velocity of the particle, we must conclude that the velocity and

---

[1] For "microscopic" particles ϑ is significantly greater since, under otherwise equal conditions, ϑ is proportional to the square of the radius of the particle.

direction of motion had changed drastically, and in a completely irregular manner, already in the extraordinarily short time ϑ.

It is therefore impossible—at least for ultramicroscopic particles—to determine $\sqrt{\overline{v^2}}$ by observation.

3. If we limit ourselves to the investigation of the paths, or—to be more precise—of the changes of position, in times τ that are substantially greater than ϑ, we will have according to the molecular theory of heat

$$\sqrt{\overline{\lambda_x^2}} = \sqrt{\tau}\sqrt{\frac{RT}{N}\cdot\frac{1}{3\pi kP}}\,,$$ [9]

where λ_x denotes the change of the particle's x-coordinate occurring during τ. As the mean velocity in the time interval τ we can define the quantity

$$\frac{\sqrt{\overline{\lambda_x^2}}}{\tau} = \frac{\omega}{\sqrt{\tau}}$$

where we put for brevity

$$\sqrt{\frac{RT}{N}\cdot\frac{1}{3\pi kP}} = \omega.$$

But this mean velocity increases as τ decreases; as long as τ is large compared with ϑ, the velocity does not approach any limiting value with decreasing τ.

Since an observer operating with certain means of observation in a certain manner can never perceive paths traveled in arbitrarily short times, a certain *mean velocity* will always appear to him as *instantaneous velocity*. [10] But it is clear that no objective property of the motion investigated corresponds to the velocity so obtained, at least if the theory corresponds to the facts. [11]

Bern, January 1907. (Received on 22 January)

Doc. 41

ON THE POSSIBILITY OF A NEW TEST OF THE RELATIVITY PRINCIPLE

by A. Einstein

[*Annalen der Physik* 23 (1907): 197-198]

In an important paper published last year, Mr. J. Stark[1] demonstrated that the moving positive ions of canal rays emit line spectra by identifying the Doppler effect and following it quantitatively. He also undertook experiments with the intention of detecting and measuring an effect of the second order (proportional to $(v/V)^2$); however, the experimental arrangement, which was not set up specifically for this purpose, was not adequate for achieving reliable results.

I will show here briefly that the principle of relativity in conjunction with the principle of the constancy of the velocity of light makes it possible to predict the above effect. As I showed in an earlier paper[2], it follows from these principles that a uniformly moving clock runs at a slower rate as judged from a "stationary" system than as judged by a co-moving observer. If ν denotes the number of the clock's strokes per unit time for the observer at rest, and ν_0 the corresponding number for the co-moving observer, then

[3] $$\frac{\nu}{\nu_0} = \sqrt{1 - \left(\frac{v}{V}\right)^2}\ ,$$

or to first approximation

$$\frac{\nu - \nu_0}{\nu_0} = -\frac{1}{2}\left(\frac{v}{V}\right)^2 .$$

The atom ion of the canal rays that emits and absorbs radiation of certain frequencies is thus to be conceived as a fast-moving clock, and the relation just indicated can therefore be applied to it.

---

[1] [1] J. Stark, *Ann. d. Phys.* 21 (1906): 401.

[2] [2] A. Einstein, *Ann. d. Phys.* 17 (1905): 903.

However, one has to take into consideration that the frequency ν_0 (for the co-moving observer) is unknown, so that the above relation is not accessible to direct experimental verification. But, it may be assumed that ν_0 is also equal to the frequency emitted or absorbed by the same ion while at rest, and this for the following reason. From the fact that one and the same line spectrum is formed under very different conditions, we conclude that the frequency ν_0 does not depend on interactions between moving ions and the stationary gas, but is a characteristic of the ion only; from this one directly concludes with the help of the principle of relativity that ν_0 must equal the frequency of radiation emitted or absorbed by an ion at rest.

The equation

$$\frac{\nu - \nu_0}{\nu_0} = -\frac{1}{2}\left[\frac{v}{V}\right]^2$$

thus gives directly the second order effect sought.

The numerical values presented by Mr. Stark for the effect are more than ten times larger than those resulting from the formula presented. It seems [4] likely to me that reliable results with regard to this problem can be expected only after it has been possible to obtain (nonluminous?) canal rays in a completely gas-free space. [5]

Bern, March 1907. (Received on 17 March 1907)

Doc. 42

CORRECTION TO MY PAPER: "PLANCK'S THEORY OF RADIATION, ETC."

by A. Einstein

[*Annalen der Physik* 22 (1907): 180-190]

In the above-cited paper, which was published in this year's January issue, I wrote: "According to Drude's investigations, these proper frequencies are to be attributed to the ponderable atoms (atom ions) themselves. The most obvious conclusion seems therefore to be to consider

exclusively the positive atom ions as the carriers of heat in solids
[1] (insulators)."

This proposition does not hold up in two respects: First, one must assume not only positively, but also negatively charged atom ions. Second—and this is the essential point—Drude's investigations do not justify the assumption that every elementary structure capable of oscillation that acts as a carrier of heat has always an electric charge. Thus, from the existence of
[2] an absorption region one can indeed deduce (within the limitations mentioned) the existence of a kind of elementary structure that makes a contribution with
[3] a characteristic temperature dependence to the specific heat; however, the converse conclusion is not valid, because most certainly there could exist uncharged heat carriers, i.e., such ones that are not observable optically. This is especially to be expected with chemically not bound atoms.

The conclusion drawn from the nature of the specific heat of diamond in the last sentence of the paper hence is also not legitimate. It should read:

"Thus, according to the theory, it is to be expected that diamond either shows an absorption maximum at $\lambda = 11\ \mu$ or that it has no optically demon-
[4] strable infrared proper frequency whatsoever."

(Received on 3 March 1907)

Erratum

Vol. 22, p. 287, line 4 from the bottom in equation (2), the letter π should be omitted.

Doc. 43
Author's abstract of lecture
"ON THE NATURE OF THE MOVEMENTS OF MICROSCOPICALLY SMALL PARTICLES SUSPENDED IN LIQUIDS"

[*Naturforschende Gesellschaft Bern. Mitteilungen* (1907)]

1038. SESSION OF 23 MARCH 1907 [1]
At 8 P.M. in the Hotel Storchen

Chairman: Mr. Ed. Fischer. Present: 20 members and guests. [2]

1. Mr. A. Einstein speaks "**On the nature of the movements of microscopically small particles suspended in liquids.**"

Microscopically small inanimate particles (e.g., with diameters of the order of magnitude 0.001 mm) suspended in liquids carry out irregular motions [3] that are more animated the smaller the diameter of the particle and the viscosity of the liquid and the higher the temperature (Brownian motion). After a brief discussion of the different attempts at explanation, a simple [4] formula for the distances covered by the particles is derived by the lecturer [5] in an elementary way with the aid of the kinetic theory of heat. (Author's abstract) [6]

For more on the topic, see: *Ann. d. Physik* 17 (1905): 549. [7]
Ann. d. Physik 19 (1906): 371.

Doc. 44

COMMENTS ON THE NOTE OF MR. PAUL EHRENFEST: "THE TRANSLATORY MOTION OF DEFORMABLE ELECTRONS AND THE AREA LAW" [1]

by A. Einstein

[*Annalen der Physik* 23 (1907): 206-208]

The article referred to above contains the following remarks:

"In the formulation in which Mr. Einstein published it, Lorentzian [2] relativistic electrodynamics is rather generally viewed as a complete system. Accordingly, it must also be able to provide purely deductively an answer to [3] the question posed by transferring Abraham's problem from the rigid electron to the deformable one: Granted that there exists a deformable electron that [4] has some nonspherical and nonellipsoidal form when at rest. According to Mr. Einstein, this electron undergoes the well-known Lorentz contraction during uniform translation. Well then, is it possible for this electron to undergo [5] force-free uniform translation in every direction, or is it not?"

Concerning this I have the following comments:

1. The principle of relativity, or, more exactly, the principle of relativity together with the principle of the constancy of velocity of light, is not to be conceived as a "complete system," in fact, not as a system at all, but merely as a heuristic principle which, when considered by itself, contains only statements about rigid bodies, clocks, and light signals. It is only by requiring relations between otherwise seemingly unrelated laws that [6] the theory of relativity provides additional statements.

For example, the theory of the motion of electrons arises in the following way. One postulates the Maxwell equations for vacuum for a system of space-time coordinates. By applying the space-time transformation derived by means of the system of relativity, one finds the transformation equations for the electric and magnetic forces. Using the latter, and applying the space-time transformation once again, one arrives at the law for the acceleration of an electron moving at arbitrary speed from the law for the acceleration of the [7] slowly moving electron (which is assumed or obtained from experience). Thus, we are not dealing here at all with a "system" in which the individual laws are implicitly contained and from which they can be found by deduction alone, but only with a principle that (similar to the second law of the theory of heat) permits the reduction of certain laws to others. [8]

2. Previously, when one did not rely on the principle of relativity, but instead tried to obtain the laws of motion of electrons by electrodynamic methods, one found it necessary to make more definite assumptions on the distribution of electricity so that the problem is not an undetermined one. The electricity was thought of as being distributed over a (rigid) framework. [9] It should be noted that the laws that govern the motion of such a structure cannot be derived from electrodynamics alone. After all, the framework is nothing other than the introduction of forces which balance the electrodynamic ones. If we view the framework as a rigid body (i.e., one not deformable by external forces), the problem of the motion of the electron can be solved deductively without arbitrariness only if the dynamics of the rigid body is known with sufficient accuracy. [10]

If the theory of relativity is correct, we are still far from the latter goal. For the time being, we only have the kinematics of parallel translation and an expression for the kinetic energy of a body in parallel translation, provided the latter does not interact with other bodies[1]; for the rest, both the dynamics and the kinematics of a rigid body have at present to be considered as unknown for the case under consideration.

Bern, 14 April 1907. (Received on 16 April 1907)

---

[1] I will soon show in an article that the latter restriction is essential. [11]

Doc. 45

ON THE INERTIA OF ENERGY REQUIRED BY THE RELATIVITY PRINCIPLE

by A. Einstein

[*Annalen der Physik* 23 (1907): 371-384]

The principle of relativity, in combination with Maxwell's equations, leads to the conclusion that the inertia of a body increases or decreases with its energy content in a completely determined way. That is to say, if one observes a body that emits a certain radiation energy simultaneously in two opposite directions, and if one examines this process from two coordinate systems which move uniformly relative to each other,[1] one of which is at rest relative to the body, and if one applies--from both coordinate systems--the energy principle to the process, one arrives at the result that to an increase in the body's energy ΔE there must always correspond an increase in the mass $\Delta E/V^2$, where V denotes the velocity of light.

The circumstance that the special case discussed there necessitates an assumption of such extraordinary generality (about the dependence of the inertia on the energy) demands that the necessity and justification of this assumption be examined in a more general way. Especially, the question arises: Do not other special cases lead to conclusions that are incompatible with the one mentioned above? A first step in this respect I took last year[2] by showing that the above assumption resolves the contradiction between electrodynamics and the principle of the constancy of the motion of the center of gravity (at least as far as the terms of first order are concerned).

The *general* answer to the question posed is not yet possible because we do not yet have a complete world view that would correspond to the principle of relativity. Rather, we must limit ourselves to the special cases that we can handle at present without arbitrariness from the standpoint of relativistic electrodynamics. We are going to consider two such cases; in the first of these, the system whose inertial mass we shall examine consists of a rigid,

---

[1] [1] A. Einstein, *Ann. d. Phys.* 18 (1905): 639.

[2] [2] A. Einstein, *Ann. d. Phys.* 20 (1906): 627.

rigidly electrified body, and in the second case it consists of a number of uniformly moving mass points which do not exert any forces on each other.

Before I start with the investigation, I must insert here a remark on the presumed range of validity of Maxwell's equations for empty space so as to meet a naturally arising objection. In earlier papers I showed that our present electromechanical world view is not suited for explaining the entropic properties of radiation and the laws governing the emission and absorption of radiation and those governing the specific heat; rather, it seems to me necessary to assume that the nature of any periodic process is such that the conversion of energy can only proceed in certain quanta of finite magnitude (light quanta), i.e., that the manifold of processes possible in reality is smaller than the manifold of processes possible according to our present theoretical views.[1] In particular, we would have to imagine that in a radiation process the instantaneous electromagnetic state in a region of space is completely determined by a *finite* number of quantities--in contrast to the vector theory of radiation. But as long as we do not possess a picture that corresponds to the requirements mentioned, it will be natural to use the current theory for all problems not concerned with entropy relations or conversions of elementary small quantities of energy without having to fear that we will thereby arrive at incorrect results. I can illustrate most graphically how I see the present situation regarding these questions with the following imaginary case.

Let us imagine that the molecular-kinetic theory of heat has not yet been propounded, but that it has been demonstrated with complete certainty that the Brownian motion (motion of particles suspended in liquids) is not due to any external supply of energy, while it is clearly recognized that these motions cannot be explained with the help of mechanics and thermodynamics. In such a situation one would rightly conclude that a radical change of theoretical principles must take place. In spite of that, nobody would shrink from applying the fundamental equations of mechanics and thermodynamics to handle problems not related to instantaneous states in small regions of space. In this sense, in my opinion, we can confidently base our considerations on Maxwell's equations.

---

[1] A. Einstein, *Ann. d. Phys.* 17 (1905): 132; 20 (1906): 199; and 22 (1907): 180. [3]

It seems to me to be in the nature of things that other authors might have already elucidated part of what I am going to say. However, bearing in mind that the problems under consideration are being treated here from a new standpoint, I felt that I should be permitted to forgo a survey of the literature (which would have been very troublesome for me), especially since there is good reason to hope that this gap will be filled by other authors, as it was kindly done by Mr. Planck and Mr. Kaufmann for my first paper on the principle of relativity. [4]

§1. *On the kinetic energy of a rigid body in uniform translation subject to external forces*

We consider a rigid body that is moving in uniform translation (velocity v) in the direction of the increasing x-coordinate of a coordinate system (x,y,z) that is assumed to be at rest. If external forces do not act upon it, then, according to the theory of relativity, its kinetic energy K_0 is given by the equation[1]

$$K_0 = \mu V^2 \left\{ \frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} - 1 \right\},$$

where μ denotes its mass (in the conventional sense) and V the velocity of light in vacuum. We now want to show that according to the theory of relativity this expression does not hold any longer if the body is acted upon by external forces that balance each other. To be able to deal with this case, we must assume that these are electrodynamic forces. We therefore imagine that the body is rigidly electrified (with continuously distributed electricity), and that an electromagnetic field of force is acting upon it. We imagine that the electric density is always very low and the field strong, so that the forces corresponding to the interactions between the body's electric masses can be neglected compared with the forces the external field exerts on

---

[5] [1] A. Einstein, *Ann. d. Phys.* 17 (1905): 917ff.

the body's electrical charges.[1] The energy ΔE transferred from the field of force to the body between the times t_0 and t_1 is given by the expression

$$\Delta E = \int_{t_0}^{t_1} dt \int vX \frac{\rho}{4\pi}\, dx dy dz \ ,$$

where the space integral is to be extended over the body and we have put

$$\rho = \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} \ .$$

Using the transformation equations given in the paper cited above[2], we transform this expression to the space-time system (ξ,η,ζ,τ), which corresponds to a coordinate system that is at rest with respect to the body and whose axes are parallel to (x,y,z). One thus obtains after a simple calculation, in a notation that corresponds exactly to that used in the paper quoted,

$$\Delta E = \iint \beta v X' \frac{\rho'}{4\pi}\, d\xi d\eta d\zeta d\tau \ ,$$

where, as there, β denotes the expression

$$\frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} \ .$$

Note that according to our assumptions the forces X' cannot be arbitrary. Rather, at all times they must be such that the body under consideration does not experience any acceleration. The necessary (but not sufficient) condition for this, according to a theorem of statics, is that, observed from a coordinate system that moves together with the body, the sum of the X-components of the forces acting upon the body always vanishes. One thus has for each τ

$$\int X'\rho' d\xi d\eta d\zeta = 0 \ .$$

---

[1] We introduce this assumption in order to be able to assume that the acting forces are not subjected to any restricting conditions due to the way they are produced.

[2] A. Einstein, *Ann. d. Phys.* 17 (1905), §§ 3 and 6. [6]

Thus, if the limits for τ in the above integral expression for ΔE were independent of ξ,η,ζ, we would have $\Delta E = 0$. However, this is not the case, for from the transformation equation

$$t = \beta\left[\tau + \frac{v}{V^2}\xi\right]$$

it follows immediately that the time limits in the moving system are

$$\tau = \frac{t_0}{\beta} - \frac{v}{V^2}\xi \quad \text{and} \quad \tau = \frac{t_1}{\beta} - \frac{v}{V^2}\xi \,.$$

We imagine that the integral in the expression for ΔE is decomposed into three parts.

The first part shall comprise the times τ between

$$\frac{t_0}{\beta} - \frac{v}{V^2}\xi \quad \text{and} \quad \frac{t_0}{\beta}\,,$$

the second part between

$$\frac{t_0}{\beta} \quad \text{and} \quad \frac{t_1}{\beta}\,,$$

and the third between

$$\frac{t_1}{\beta} \quad \text{and} \quad \frac{t_1}{\beta} - \frac{v}{V^2}\xi \,.$$

The second part vanishes because its time limits are independent of ξ,η,ζ. The first and third parts have a definite value only if the assumption is made that the forces acting on the body are independent of time close to the times $t = t_0$ and $t = t_1$, such that the electric force X' is independent of time for all points of the rigid body between the times

$$\tau = \frac{t_0}{\beta} - \frac{v}{V^2}\xi \quad \text{and} \quad \tau = \frac{t_0}{\beta}\,,$$

and between

$$\tau = \frac{t_1}{\beta} \quad \text{and} \quad \tau = \frac{t_1}{\beta} - \frac{v}{V^2}\xi\,,$$

respectively. If the X' present during these two time intervals are called X_1' and X_2', respectively, one obtains

$$\Delta E = -\int \frac{v^2}{V^2} \beta \frac{\xi X_1' \rho'}{4\pi} d\xi d\eta d\zeta + \int \frac{v^2}{V^2} \beta \frac{\xi X_0' \rho'}{4\pi} d\xi d\eta d\zeta \ .$$

If one assumes further that no forces act on the body at the start $(t = t_0)$, then the second of these integrals vanishes. Taking into account that

$$\frac{X_1' \rho'}{4\pi} d\xi d\eta d\zeta$$

is the ξ-component K_ξ of the ponderomotive force acting on the space element, one gets

$$\Delta E = -\frac{\left(\frac{v}{V}\right)^2}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} \sum (\xi K_\xi) \ ,$$

where the summation is to be extended over all mass elements of the body.

We thus get the following strange result. If a rigid body on which originally no forces are acting is subjected to the influence of forces that do not impart acceleration to the body, then these forces—observed from a coordinate system that is moving relative to the body—perform an amount of work ΔE on the body that depends only on the final distribution of forces and the translation velocity. In accordance with the energy principle, from this it follows immediately that the kinetic energy of a rigid body subjected to forces is larger by ΔE than the kinetic energy of the same body moving at the same velocity but not subjected to any forces.

§2. *On the inertia of an electrically charged rigid body*

We again consider a rigid, rigidly electrified body in uniform translation (velocity v) in the direction of the increasing x-coordinate of a "stationary" coordinate system. An external electromagnetic field of force shall not be present. This time, however, we shall take into account the electromagnetic field produced by the electric masses of the body. First, we calculate the electromagnetic energy

$$E_e = \frac{1}{8\pi}\int (X^2 + Y^2 + Z^2 + L^2 + M^2 + N^2)dxdydz \ .$$

To this end we use the transformation equations contained in the repeatedly cited paper, and transform the above expression by introducing under the integral the quantities that referred to a coordinate system moving with the body. We then obtain

$$E_e = \frac{1}{8\pi}\int \frac{1}{\beta}\left[X'^2 + \frac{1 + (\frac{v}{V})^2}{1 - (\frac{v}{V})^2}(Y'^2 + Z'^2)\right]d\xi d\eta d\zeta \ .$$

It should be noted that the value of this expression depends on the orientation of the rigid body relative to the direction of motion. Hence, if the total kinetic energy of the electrified body consisted exclusively of the kinetic energy K_0 of the body due to its ponderable mass and of the excess of the electromagnetic energy of the moving body over the electrostatic energy of the body when at rest, we would have arrived at a contradiction, as we can easily see from the following.

We imagine that the body under consideration rotates infinitely slowly relative to the coordinate system moving along with it, with no external influences taking place during this motion. It is clear that this motion must be possible without application of any force, because according to the principle of relativity the body's laws of motion relative to the system moving along with it are the same as the laws of motion with respect to a "stationary" system. We now observe the uniformly moving and infinitely slowly rotating body from the "stationary" system. Since the rotation is supposed to be infinitely slow, it does not contribute anything to the kinetic energy. The expression for the kinetic energy in the case under consideration is therefore the same as it would be if no rotation but only uniform parallel translation were to take place. However, since in the course of the motion the body takes up different (arbitrary) positions, and the energy principle must hold throughout the motion, it is clear that the kinetic energy of an electrified body in translational motion cannot possibly depend on its orientation.

This contradiction is resolved by the results of the previous section, i.e., the kinetic energy of the body under consideration cannot be calculated like that of a rigid body upon which no forces are acting. On the contrary, in accordance with §1, we must take into account that our rigid body is subjected to forces caused by the interaction between the electric masses. Thus, if we denote by K_0 the kinetic energy in the absence of electric charges, we obtain for the body's total kinetic energy K the expression

$$K = K_0 + \Delta E + (E_e - E_s) \ ,$$

where E_s denotes the electrostatic energy of the body in the state of rest. In our case we have

$$\Delta E = -\frac{v^2}{V^2}\beta\frac{1}{4\pi}\int \xi X'\left[\frac{\partial X'}{\partial \xi} + \frac{\partial Y'}{\partial \eta} + \frac{\partial Z'}{\partial \zeta}\right]d\xi d\eta d\zeta \ ,$$

from which one obtains by integration by parts, taking into consideration that X', Y', Z' can be derived from a potential,

$$\Delta E = \frac{v^2}{V^2}\beta\frac{1}{8\pi}\int \left[X'^2 - Y'^2 - Z'^2\right]d\xi d\eta d\zeta \ .$$

If one takes into account the expressions for K_0 and β given in §1, one obtains the following expression for the kinetic energy of the electrified rigid body:

$$K = \left[\mu + \frac{E_s}{V^2}\right] \cdot V^2\left[\frac{1}{\sqrt{1-\left(\frac{v}{V}\right)^2}} - 1\right] .$$

This expression is, as it must be, independent of the body's orientation relative to the direction of translation. If one compares the expression for K with that for the energy K_0 of a body not charged electrically,

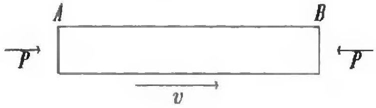
$$K_0 = \mu V^2\left[\frac{1}{\sqrt{1-\left(\frac{v}{V}\right)^2}} - 1\right] ,$$

one realizes that the electrostatically charged body has an inertial mass that exceeds that of the uncharged body by the electrostatic energy divided by the square of the velocity of light. The law of the inertia of energy is thus confirmed by our result in the special case considered.

§3. *Remarks concerning the dynamics of the rigid body*

From the foregoing it might seem that we are no longer far from the goal of constructing a dynamics of the parallel translation of the rigid body that [7] would conform to the principle of relativity. However, one must remember that the investigation carried out in §1 yielded the energy of a rigid body subjected to forces only for the case that these forces are constant in time. If at the time t_1 the forces X' depend on the time, then the work ΔE, and thus also the energy of the rigid body, proves to be dependent not only on those forces that occur at *one* particular time.

To illustrate the difficulty involved as drastically as possible, let us imagine the following simple special case. We consider a rigid rod AB which shall be at rest relative to a coordinate system (ξ,η,ζ), with the rod axis resting in the ζ-axis. At a certain time τ_0 let equal but opposite forces P act on the rod ends for a very short time, while at all other times the rod is not subjected to forces. It is obvious that the above action on the rod at time τ_0 does *not* produce any motion of the rod. We now observe the very same process from a coordinate system whose axes are parallel to those of the system used earlier, relative to which the rod moves in the direction A-B with velocity v. However, viewed from this coordinate system, the impulses in A and B do not act simultaneously; rather, the impulse in B is delayed by $\ell\beta(v/V^2)$ time units with regard to the impulse in A, where ℓ denotes the length of the rod (measured at rest). Thus we arrived at the following odd-looking result. On the moving rod AB, an impulse acts first in



A and some time thereafter an opposite one in B. The two impulses compensate each other so that they do not modify the motion of the rod. The case looks even more odd if we ask about the energy at a time when the impulse in A had already ended while that in B had not yet begun. The impulse in A had transferred work to the rod (since the rod was in motion); hence the energy of the rod had to increase by this work. Yet no change has occurred either in the velocity of the rod or in any other related quantity on which the energy function might be made to depend. Thus there appears to be a violation of the energy principle.

This difficulty has a very simple solution *in principle*. By implicitly assuming that we can completely determine the momentary state of the rod by the forces acting on the rod and by the rod's velocity at that moment, we assume that an increase in the body's velocity is produced instantaneously by a velocity-producing force acting somewhere on the body, i.e., that the spreading of the force exerted on one point of the body over the whole body does not require time. As we are going to show, such an assumption is not compatible with the principle of relativity. We are therefore obviously forced to postulate in our case that the effect of the impulse in A is associated with a change of state of unknown quality in the body, which spreads throughout it with finite velocity and produces an acceleration of the body in a short time unless this effect is compensated by the effects of some other forces acting upon the body within that time. Hence, if relativistic electrodynamics is correct, we are still far from having a dynamics of the parallel translation of the rigid body. [8]

We will now show that not only the assumption of an *instantaneous* spread of some effect, but also, more generally, any assumption of the spreading of an effect with a velocity greater than the velocity of light is incompatible with the theory of relativity. [9]

Consider a material strip extending along the x-axis of a coordinate system (x,y,z), relative to which a certain effect shall propagate with velocity W, and let there be at $x = 0$ (point A) as well as at $x = +\ell$ (point B) an observer who is at rest relative to the coordinate system (x,y,z). By means of the above effect, the observer in A sends a signal to the observer in B through the material strip, which is not at rest but is moving in the negative x-direction with the velocity v $(< V)$. It follows from §5 (*loc. cit.*) that the signal will then be transmitted from A to B with the velocity

$$\frac{W - v}{1 - \frac{Wv}{V^2}} .$$

The time T that elapses between the signal emission in A and signal reception in B is thus

$$T = \ell\,\frac{1 - \frac{Wv}{V^2}}{W - v} .$$

The velocity v can assume any value smaller than V. Hence, if $W > V$, as we have assumed, then v can always be chosen such that $T < 0$. This result signifies that we would have to consider as possible a transfer mechanism whose use would produce an effect which *precedes* the cause (accompanied by an act of will, for example). Even though, in my opinion, this result does not contain a contradiction from a purely logical point of view, it conflicts so absolutely with the character of all our experience, that the impossibility of [10] the assumption $W > V$ is sufficiently proved by this result.

§4. *On the energy of a system consisting of a number of mass points moving force-free*

If one takes a look at the expression for the kinetic energy k of a mass point (μ) moving with the velocity v,

$$k = \mu V^2 \left\{ \frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} - 1 \right\} ,$$

one notices that this expression has the form of a difference; i.e. we have

$$k = \left| \mu V^2 \frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} \right|_{v = 0}^{v = v} .$$

If one does not ask about the kinetic energy in particular, but simply about the energy ϵ of the moving mass point, then $\epsilon = k + \text{const}$. While it is most convenient to set the arbitrary constant in this equation equal to zero in classical mechanics, the simplest expression for ϵ in relativistic mechanics is obtained by choosing the zero point such that the energy ϵ_0 for the stationary mass point equals μV^2.[1] One then obtains

$$\epsilon = \mu V^2 \frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} .$$

We will henceforth adhere to this choice of the zero point of the energy.

We now introduce again the two coordinate systems (x,y,z) and (ξ,η,ζ) that are always moving relative to each other. Let a mass point μ move relative to (ξ,η,ζ) with a velocity w in a direction that forms the angle φ with the positive ξ-axis. The energy ϵ of the mass point relative to the system (x,y,z) can easily be determined using the relations derived in §5 (*loc. cit.*). One obtains

$$\epsilon = \mu V^2 \frac{1 + \frac{vw \cos \varphi}{V^2}}{\sqrt{1 - \left(\frac{v^2}{V^2}\right)} \sqrt{1 - \left(\frac{w^2}{V^2}\right)}} .$$

If several mass points are present that have different masses, velocities, and directions of motions, we obtain for their total energy E the expression

$$E = \frac{1}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} \left\{ \sum \mu V^2 \cdot \frac{1}{\sqrt{1 - \left(\frac{w}{V}\right)^2}} \right\} + \frac{v}{\sqrt{1 - \left(\frac{v}{V}\right)^2}} \left\{ \sum \frac{\mu w \cos \varphi}{\sqrt{1 - \left(\frac{w}{V}\right)^2}} \right\}$$

Until now we have not stipulated anything about the state of motion of the system (ξ,η,ζ) relative to the moving masses. We can and will now stipulate

---

[1] One should note that the simplifying stipulation $\mu V^2 = \epsilon_0$ is also the expression of the principle of the equivalence of mass and energy, and that in the case of the electrified body ϵ_0 is nothing other than its electrostatic energy. [11]

the following conditions that uniquely determine the state of motion of (ξ,η,ζ):

$$\sum \frac{\mu w_\xi}{\sqrt{1-\left(\frac{w}{V}\right)^2}} = 0 , \quad \sum \frac{\mu w_\eta}{\sqrt{1-\left(\frac{w}{V}\right)^2}} = 0 , \quad \sum \frac{\mu w_\zeta}{\sqrt{1-\left(\frac{w}{V}\right)^2}} = 0 ,$$

where w_ξ, w_η, w_ζ denote the components of w. In classical mechanics this stipulation corresponds to the condition that the momentum of motion of the mass system relative to (ξ,η,ζ) vanishes. We then obtain

$$E = \left[\sum \mu V^2 \cdot \frac{1}{\sqrt{1-\left(\frac{w}{V}\right)^2}}\right] \cdot \frac{1}{\sqrt{1-\left(\frac{v}{V}\right)^2}} ,$$

or, by introducing the energy E_0 of the system relative to the system (ξ,η,ζ),

$$E = \frac{E_0}{V^2} \cdot V^2 \frac{1}{\sqrt{1-\left(\frac{v}{V}\right)^2}} .$$

If this expression is compared with that for the energy of a mass point moving with the velocity v,

$$\epsilon = \mu V^2 \frac{1}{\sqrt{1-\left(\frac{v}{V}\right)^2}} ,$$

one obtains the following result: With respect to the dependence of the energy on the state of motion of the coordinate system to which the processes are referred, a system of uniformly moving mass points can be replaced by a single mass point having the mass $\mu = E_0/V^2$.

Thus, a system of moving mass points—taken as a whole—has the more inertia the faster the mass points move relative to each other. The dependence is again given by the law cited in the Introduction.

Bern, May 1907. (Received on 14 May 1907)

Doc. 46

Review of J. J. WEYRAUCH, *An Outline of the Theory of Heat. With Numerous Examples and Applications. Part 2* (*Grundriss der Wärmetheorie. Mit zahlreichen Beispielen und Anwendungen.* Stuttgart: K. Wittwer, 1907. 412 pp. 16.00 mark.) [1]

[*Beiblätter zu den Annalen der Physik* 31 (1907): 777]

The second volume considered here deals with saturated and superheated vapors, with steam engines, aerostatics, aerodynamics (motion in channels, efflux from container orifices), and with solid bodies; together with the first volume, the present volume is superbly suited to introduce the person familiar with the elements of the differential and integral calculus to the theory of heat in a way that is useful for the treatment of engineering problems (especially those concerned with engines). 150 problems and 250 numerical examples, interlaced with the text and mostly selected corresponding [2] to conditions found in practice, give the student the opportunity for ample exercise, so that he may acquire relatively easily the proficiency needed for the reading of engineering papers and for the calculation of special problems. In addition, the book supplies him with the necessary empirical data as well as with numerous references to the (mostly engineering) literature. The editing of the book is simple and clear, the application of mathematical symbols is consistent, the table of contents is clearly arranged and complete, so that one can get informed about specific questions without losing much time.

Doc. 47

ON THE RELATIVITY PRINCIPLE AND THE CONCLUSIONS DRAWN FROM IT

by A. Einstein

[*Jahrbuch der Radioaktivität und Elektronik* 4 (1907): 411-462]

Newton's equations of motion retain their form when one transforms to a new system of coordinates that is in uniform translational motion relative to the system used originally according to the equations

$$x' = x - vt$$

$$x' = y$$

$$z' = z \ .$$

As long as one believed that all of physics can be founded on Newton's equations of motion, one therefore could not doubt that the laws of nature are the same without regard to which of the coordinate systems moving uniformly (without acceleration) relative to each other they are referred. However, this independence from the state of motion of the system of coordinates used, which we will call "the principle of relativity," seemed to have been suddenly called into question by the brilliant confirmations of H. A. Lorentz's electrodynamics of moving bodies.[1] That theory is built on the presupposition of a resting, immovable, luminiferous ether; its basic equations are not such that they transform to equations of the same form when the above transformation equations are applied.

After the acceptance of that theory, one had to expect that one would succeed in demonstrating an effect of the terrestrial motion relative to the luminiferous ether on optical phenomena. It is true that in the study cited
[2] Lorentz proved that in optical experiments, as a consequence of his basic assumptions, an effect of that relative motion on the ray path is not to be expected as long as the calculation is limited to terms in which the ratio

---

[1] [1] H. A. Lorentz, *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern.* [Attempt at a theory of electric and optical phenomena in moving bodies] Leiden, 1895. Reprinted Leipzig, 1906.

v/c of the relative velocity to the velocity of light in vacuum appears in the first power. But the negative result of Michelson and Morley's experiment[1] showed that in a particular case an effect of the second order (proportional to v^2/c^2) was not present either, even though it should have shown up in the experiment according to the fundamentals of the Lorentz theory.

It is well known that this contradiction between theory and experiment was formally removed by the postulate of H. A. Lorentz and FitzGerald, [4] according to which moving bodies experience a certain contraction in the direction of their motion. However, this ad hoc postulate seemed to be only an artificial means of saving the theory: Michelson and Morley's experiment had actually shown that phenomena agree with the principle of relativity even where this was not to be expected from the Lorentz theory. It seemed therefore as if Lorentz's theory should be abandoned and replaced by a theory whose foundations correspond to the principle of relativity, because such a theory would readily predict the negative result of the Michelson and Morley experiment. [5]

Surprisingly, however, it turned out that a sufficiently sharpened conception of time was all that was needed to overcome the difficulty discussed. One had only to realize that an auxiliary quantity introduced by H. A. Lorentz and named by him "local time" could be defined as "time" in [6] general. If one adheres to this definition of time, the basic equations of Lorentz's theory correspond to the principle of relativity, provided that the above transformation equations are replaced by ones that correspond to the new conception of time. H. A. Lorentz's and FitzGerald's hypothesis appears then as a compelling consequence of the theory. Only the conception of a luminiferous ether as the carrier of the electric and magnetic forces does not fit into the theory described here; for electromagnetic forces appear here not as states of some substance, but rather as independently existing things that are similar to ponderable matter and share with it the feature of inertia. [7]

The following is an attempt to summarize the studies that have resulted to date from the merger of the H. A. Lorentz theory and the principle of relativity.

---

[1] A. A. Michelson and E. W. Morley, *Amer. J. of Science* 34, (1887): 333. [3]

The first two parts of the paper deal with the kinematic foundations as well as with their application to the fundamental equations of the Maxwell-Lorentz theory, and are based on the studies[1] by H. A. Lorentz (*Versl. Kon.*
[9] *Akad. v. Wet., Amsterdam* (1904)) and A. Einstein (*Ann. d. Phys.* 16 (1905)).

In the first section, in which only the kinematic foundations of the theory are applied, I also discuss some optical problems (Doppler's principle, aberration, dragging of light by moving bodies); I was made aware of the possibility of such a mode of treatment by an oral communication and a paper
[10] by Mr. M. Laue (*Ann. d. Phys.* 23 (1907): 989), as well as a paper (though in
[11] need of correction) by Mr. J. Laub (*Ann. d. Phys.* 32 (1907)).

In the third part I develop the dynamics of the material point (electron). In the derivation of the equations of motion I used the same method as
[12] [13] in my paper cited earlier. Force is defined as in Planck's study. The reformulations of the equations of motion of material points, which so clearly demonstrate the analogy between these equations of motion and those of classical mechanics, are also taken from that study.

The fourth part deals with the general inferences regarding the energy and momentum of physical systems to which one is led by the theory of relativity. These have been developed in the original studies,

> [14] A. Einstein, *Ann. d. Phys.* 18 (1905): 639 and *Ann. d. Phys.* 23 (1907): 371, as well as M. Planck, *Sitzungsber. d. Kgl. Preuss. Akad. d. Wissensch.* XXIX (1907),

but are here derived in a new way, which, it seems to me, shows especially clearly the relationship between the above application and the foundations of the theory. I also discuss here the dependence of entropy and temperature on the state of motion; as far as entropy is concerned, I kept completely to the Planck study cited, and the temperature of moving bodies I defined as did Mr. Mosengeil in his study on moving black-body radiation.[2]

The most important result of the fourth part is that concerning the inertial mass of the energy. This result suggests the question whether energy also possesses *heavy* (gravitational) mass. A further question suggesting

---

[8] [1] E. Cohn's studies on the subject are also pertinent, but I did not make use of them here.

[15] [2] Kurd von Mosengeil, *Ann. d. Phys.* 22 (1907): 867.

itself is whether the principle of relativity is limited to *nonaccelerated* moving systems. In order not to leave this question totally undiscussed, I added to the present paper a fifth part that contains a novel consideration, based on the principle of relativity, on acceleration and gravitation.

I. KINEMATIC PART

§1. *Principle of constancy of the velocity of light.*
Definition of time. Principle of relativity.

To be able to describe a physical process, we must be able to evaluate the changes taking place at the individual points of the space as functions of position and time.

To determine the position of a process of infinitesimally short duration that occurs in a space element (point event) we need a Cartesian system of coordinates, i.e., three mutually perpendicular rigid rods rigidly connected with each other, and a rigid unit measuring rod.[1] Geometry permits us to determine the position of a point, i.e., the location of a point event, by means of three numbers (coordinates x, y, z).[2] To evaluate the time of a point event, we use a clock that is at rest relative to the coordinate system and in whose immediate vicinity the point event takes place. The time of the point event is defined by the simultaneous clock reading.

Imagine that clocks at rest with respect to the coordinate system are arranged at many points. Let all these clocks be equivalent, i.e., the difference between the readings of two such clocks shall remain unchanged if they are arranged next to each other. If these clocks are imagined to be set in some manner, then the totality of the clocks, provided they are arranged sufficiently closely, will permit the temporal evaluation of any point event, say by using the nearest clock.

---

[1] Instead of speaking of "rigid" bodies, we could equally well speak, here, as well as further on, of solid bodies not subjected to deforming forces. [16]

[2] For this one also needs auxiliary rods (rulers, compasses).

However, the totality of these clock readings does not yet give us the "time" as we need it for physical purposes. For this we also need a rule according to which these clocks will be set relative to each other.

We now assume *that the clocks can be adjusted in such a way that the propagation velocity of every light ray in vacuum—measured by means of these clocks—becomes everywhere equal to a universal constant* c, provided that the coordinate system is not accelerated. If A and B are two points at rest relative to the coordinate system, which are equipped with clocks and are separated by a distance r, while t_A is the reading of the clock in A at the moment when a ray of light propagating through the vacuum in the direction AB reaches the point A, and t_B is the reading of the clock at B at the moment the ray reaches B, then we should always have

$$\frac{r}{t_B - t_A} = c ,$$

whatever the motion of the light source emitting the light ray or the motion of other bodies may be.

It is by no means self-evident that the assumption made here, which we will call "the principle of the constancy of the velocity of light," is actually realized in nature, but—at least for a coordinate system in a certain state of motion—it is made plausible by the confirmation of the Lorentz theory[1], which is based on the assumption of an ether that is absolutely at rest, through experiment[2].

The aggregate of the readings of all clocks synchronized according to the above, which may be imagined as being arranged in the individual points of space at rest with respect to the coordinate system, we call the time belonging to the coordinate system used, or, in short, the time of that system.

The coordinate system used, together with the unit measuring rod and the clocks that serve for the determination of the time of the system, we call "reference system S." Suppose that the physical laws are ascertained with

---

[17] [1] H. A. Lorentz, *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern* [Attempt at a theory of electrical and optical phenomena in moving bodies]. Leiden, 1895.

[18] [2] It is of special relevance that this theory furnished the drag coefficient (Fizeau's experiment) in accordance with experience.

respect to the reference system S that is at first at rest relative to the sun. Let then the reference system S be accelerated by some external cause for awhile, and, finally, let it return to a nonaccelerated state. What will the physical laws look like when the processes are referred to the system S that is now in another state of motion?

We now make the simplest possible assumption, which is also suggested by the Michelson and Morley experiment: *The physical laws are independent of the state of motion of the reference system, at least if the system is not accelerated.* [19]

In the ensuing considerations, we will base ourselves on this assumption, which we call "the principle of relativity," as well as on the principle of the constancy of the velocity of light set forth above.

§2. *General remarks concerning space and time*

1. We consider a number of rigid bodies in nonaccelerated motion with equal velocities (i.e., at rest relative to each other). In accordance with the principle of relativity, we conclude that the laws according to which these bodies can be grouped in space relative to each other do not change with the change of these bodies' common state of motion. From this it follows that the laws of geometry determine the possible arrangements of rigid bodies in nonaccelerated motion always in the same way, independent of their common state of motion. Assertions about the shape of a body in nonaccelerated motion therefore have a direct meaning. The shape of a body in the sense indicated we will call its "geometric shape." The latter obviously does not depend on the state of motion of a reference system. [20]

2. According to the definition of time given in §1, a statement on time has a meaning only with reference to a reference system in a specific state of motion. It may therefore be surmised (and will be shown in what follows) that two spatially distant point events that are simultaneous with respect to a reference system S are in general not simultaneous with respect to a reference system S' whose state of motion is different.

3. Suppose a body consisting of material points P moves in some manner relative to a reference system S. At time t of S, each material point P

occupies a certain position in S, i.e., coincides with a certain point Π that is at rest relative to S. The totality of positions of points Π relative to the coordinate system S we call position, and the totality of the interrelations of positions of points P we call the kinematic shape of the body with respect to S for the time t. If the body is at rest relative to S, its kinematic shape is identical with the geometric one.

It is clear that observers who are at rest relative to a reference system S can ascertain only the *kinematic* shape with respect to S of a body that is in motion relative to S, but not its geometric shape.

In the following, we will usually not distinguish explicitly between geometric and kinematic shape; a statement of geometric nature refers to kinematic or geometric shape, respectively, depending on whether the latter refers to a reference system S or not.

[21]

§3. *Transformation of coordinates and time*

Let S and S' be equivalent reference systems, i.e., these systems shall have unit measuring rods of the same length and clocks running at the same rate when these objects are compared with each other in a state of relative rest. It is then obvious that all physical laws that hold with respect to S will hold in exactly the same form for S' too, if S and S' are at rest relative to each other. The principle of relativity requires such total equivalence also if S' is in uniform translational motion with respect to S. Hence, specifically, the velocity of light in vacuum must have the same numerical value with respect to both systems.

Let a point event be determined by the variables x, y, z, t with respect to S, and by the variables x', y', z', t' with respect to S', where S and S' are moving without acceleration and relative to each other. We seek the equations that relate the former to the latter variables.

Right off, we can state about these equations that they must be linear with respect to these variables because this is required by the homogeneity properties of space and time. Specifically, from this it follows that the coordinate planes of S' are uniformly moving planes with respect to S; yet in general these planes will not be perpendicular to each other. However, if we choose the position of the x'-axis in such a way that it has, with reference to S, the same direction as the translational motion of S' has with reference to S, then it follows for reasons of symmetry that the S-referred coordinate planes of S' must be mutually perpendicular. We can and will choose the positions of the two coordinate systems in such a way that the x-axis of S and the x'-axis of S' coincide at all times, and that the S-referred y'-axis of S' be parallel to the y-axis of S. Further, we shall choose the instant at which the coordinate origins coincide as the starting time in both systems; the linear transformation equations sought are then homogeneous.

From the now known position of the coordinate planes of S' relative to S, we immediately conclude that the following pairs of equations are equivalent:

$$\begin{aligned} x' &= 0 \quad \text{and} \quad x - vt = 0 \\ y' &= 0 \quad \text{and} \quad y = 0 \\ z' &= 0 \quad \text{and} \quad z = 0 \end{aligned}$$

Three of the transformation equations sought thus have the form:

$$\begin{aligned} x' &= a(x - vt) \\ y' &= by \\ z' &= cz \ . \end{aligned}$$

Since the propagation velocity of light in empty space is c with respect to both reference systems, the two equations

$$x^2 + y^2 + z^2 = c^2t^2$$

and

$$x'^2 + y'^2 + z'^2 = c^2t'^2$$

must be equivalent. From this and the expressions for x', y', z' just found we conclude after a simple calculation that the transformation equations must be of the form

$$\begin{aligned} t' &= \varphi(v)\cdot\beta\cdot\left[t - \frac{v}{c^2}x\right] \\ x' &= \varphi(v)\cdot\beta\cdot(x - vt) \\ y' &= \varphi(v)\cdot y \\ z' &= \varphi(v)\cdot z \ , \end{aligned}$$

where

$$\beta = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \ .$$

Now we will determine the function of v, which has not yet been determined. If we introduce a third system, S'', which is equivalent to S and S', is moving with the velocity $-v$ relative to S', and is oriented relative to S' in the same way S' is oriented relative to S, we obtain, by twofold application of the equations we have just found,

$$\begin{aligned} t'' &= \varphi(v)\cdot\varphi(-v)\cdot t \\ x'' &= \varphi(v)\cdot\varphi(-v)\cdot x \\ y'' &= \varphi(v)\cdot\varphi(-v)\cdot y \\ z'' &= \varphi(v)\cdot\varphi(-v)\cdot z \ . \end{aligned}$$

Since the coordinate origins of S and S'' coincide permanently, the axes have identical directions and the systems are "equivalent," this substitution is the identity,[1] so that

$$\varphi(v)\cdot\varphi(-v) = 1 \ .$$

Further, since the relation between y and y' cannot depend on the sign of v, we have

$$\varphi(v) = \varphi(-v) \ .$$

Thus,[2] $\varphi(v) = 1$, and the transformation equations read

---

[1] This conclusion is based on the physical assumption that the length of a measuring rod or the rate of a clock do not undergo any permanent changes if these objects are set in motion and then brought to rest again.

[2] $\varphi(v) = -1$ is obviously out of the question.

$$\left.\begin{aligned} t' &= \beta\left[t - \frac{v}{c^2}x\right] \\ x' &= \beta(x - vt) \\ y' &= y \\ z' &= z \quad , \end{aligned}\right\} \qquad (1)$$

where

$$\beta = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \; .$$

If we solve equations (1) for x, y, z, and t, we obtain the same equations, except that the "primed" quantities are replaced by the corresponding "unprimed" ones, and vice versa, and that v is replaced by $-v$. This also follows directly from the principle of relativity and from the fact that, relative to S', S performs a parallel translation with velocity $-v$ in the direction of the X'-axis. [22]

In general, according to the principle of relativity each correct relation between "primed" (defined with respect to S') and "unprimed" (defined with respect to S) quantities or between quantities of only one of these kinds yields again a correct relation if the unprimed symbols are replaced by the corresponding primed symbols, or vice versa, and if v is replaced by $-v$.

§4. *Inferences from the transformation equations concerning rigid bodies and clocks*

1. Let a body be at rest relative to S'. Let x_1', y_1', z_1' and x_2', y_2', z_2' be the coordinates of two material points of the body with respect to S'. In accordance with the transformation equations just derived, the following relations hold between the x_1, y_1, z_1 and x_2, y_2, z_2 coordinates of these points relative to the reference system S at all times t of S:

$$\left.\begin{aligned} x_2 - x_1 &= \sqrt{1 - \frac{v^2}{c^2}}\,(x_2' - x_1') \\ y_2 - y_1 &= y_2' - y_1' \\ z_2 - z_1 &= z_2' - z_1' \ . \end{aligned}\right\} \qquad (2)$$

The kinematic shape of a body undergoing uniform translational motion thus always depends on its velocity relative to the reference system; actually, the body's kinematic shape differs from its geometric shape only by a contraction in the direction of the relative motion in the ratio of $1 : \sqrt{1 - \frac{v^2}{c^2}}$. A relative motion of reference systems with superlightvelocity is not compatible with our principles.

2. In the coordinate origin of S' let there be set up a clock at rest which runs ν_0 times faster than the clocks used for measuring the time in S and S', i.e., this clock shall complete ν_0 periods during the time a clock at rest relative to it, of the type used for measuring time in S and S', increases its reading by one unit. How fast does the first clock run as observed from system S?

The clock considered completes one period in the time epochs $t_n' = \frac{n}{\nu_0}$, where n runs through the integers, and $x' = 0$ for the clock at all times. Using the first two transformation equations, one obtains for the time epochs t_n in which the clock, as viewed from S, completes one period:

$$t_n = \beta t_n' = \frac{\beta}{\nu_0} n \ .$$

Thus, observed from the system S, the clock completes $\nu = \frac{\nu_0}{\beta} = \nu_0\sqrt{1 - \frac{v^2}{c^2}}$ periods per unit time; or: the rate of a clock moving uniformly with velocity v relative to a reference system is slower in the ratio $1 : \sqrt{1 - \frac{v^2}{c^2}}$, as observed from this system, than that of the same clock when at rest relative to that system.

The formula $\nu = \nu_0\sqrt{1 - \frac{v^2}{c^2}}$ permits a very interesting application. Mr. J. Stark showed last year[1] that the ions constituting canal rays emit line spectra by observing a shift in spectral lines which he interpreted as a Doppler effect.

Since the oscillation process that corresponds to a spectral line is to be considered an intra-atomic process whose frequency is determined by the ion alone, we may consider such an ion as a clock of a certain frequency ν_0, which can be determined, for example, by investigating the light emitted by identically constituted ions which are at rest relative to the observer. The above consideration shows, then, that the effect of motion on the light frequency that is to be ascertained by the observer is not completely given by the Doppler effect. The motion also reduces the (apparent) proper frequency of the emitting ions in accordance with the relation given above.[2] [24]

§5. *The addition theorem of velocities*

Let a point move uniformly relative to the system S' according to the equations

$$x' = u'_x t'$$

$$y' = u'_y t'$$

$$z' = u'_z t' .$$

If x', y', z', t' are replaced by their expressions in x, y, z, t with the help of transformation equations (1), one obtains x, y, z as functions of t, and thus also the point's velocity components w_x, w_y, w_z with respect to S. We thus get

---

[1] J. Stark, *Ann. d. Phys.* 21 (1906): 401. [23]

[2] Cf. §6, equation (4a).

[25]

$$\left.\begin{aligned} u_x &= \frac{u'_x + v}{1 + \frac{vu'_x}{c^2}} \\ u_y &= \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{vu'_x}{c^2}}\, u'_y \\ u_z &= \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{vu'_x}{c^2}}\, u'_z\,. \end{aligned}\right\} \qquad (3)$$

The law of the parallelogram of velocities thus holds only in first approximation. If we set

$$u^2 = u_x^2 + u_y^2 + u_z^2$$

$$u'^2 = u'^2_x + u'^2_y + u'^2_z$$

and denote by α the angle between the x'-axis (v) and the point's direction of motion relative to S' (w'), we will have

[26]

$$u = \frac{\sqrt{(v^2 + u'^2 + 2vu' \cos \alpha) - \left[\frac{vu' \sin \alpha}{c^2}\right]^2}}{1 + \frac{vu' \cos \alpha}{c^2}}\,.$$

If the two velocities (v and u') have the same direction, we have

$$u = \frac{v + u'}{1 + \frac{vu'}{c^2}}\,.$$

It follows from this equation that the addition of two velocities smaller than c always results in a velocity smaller than c; i.e., if one sets $v = c - k$, $u' = c - \lambda$, where k and λ are positive and smaller than c, then

$$u = c\,\frac{2c - k - \lambda}{2c - k - \lambda + \frac{k\lambda}{c}} < c\,.$$

It follows further that the addition of the light velocity c and a "sublightvelocity" yields again the light velocity c.

The addition theorem of velocities also yields the interesting conclusion that there cannot exist an effect that can be used for arbitrary signaling and that is propagated faster than light in vacuum. For example, [27] let there be a material strip stretched along the x-axis of S, relative to which a certain effect (viewed from the material strip) propagates with velocity W, and let there be two observers, one in the point $x = 0$ (point A) and one in the point $x = \lambda$ (point B) of the x-axis, who are at rest relative to S. Let the observer in A send a signal by means of the [28] above-mentioned effect to the observer in B through the material strip, which shall not be at rest but shall be moving in the *negative* x-direction with velocity v $(< c)$. As a consequence of the first of equations (3), the signal will then be transmitted from A to B with velocity $\frac{W - v}{1 - \frac{Wv}{c^2}}$. The time T necessary for this is then

$$T = \ell \frac{1 - \frac{Wv}{c^2}}{W - v} .$$

The velocity v can assume any value smaller than c. Hence, if, as we have assumed, $W > c$, one can always choose v such that $T < 0$. This result means that we would have to consider as possible a transfer mechanism whereby the achieved effect would precede the cause. Even though this result, in my opinion, does not contain any contradiction from a purely logical point of view, it conflicts with the character of all our experience to such an extent that this seems sufficient to prove the impossibility of the assumption $W > c$.

§6. *Application of the transformation equations to some problems in optics* [29]

Suppose the light vector of a plane light wave propagated in vacuum is proportional to

$$\sin \omega\left[t - \frac{\ell x + my + nz}{c}\right]$$ [30]

with respect to the system S, and to

$$\sin \omega' \left[t' - \frac{\ell' x + m' y' + n' z'}{c} \right]$$

with respect to S'. The transformation equations developed in §3 require the following relations between the quantities ω, ℓ, m, n and ω', ℓ', m', n':

$$\left.\begin{aligned}
\omega' &= \omega\beta\left[1 - \ell\frac{v}{c}\right] \\
\ell' &= \frac{\ell - \frac{v}{c}}{1 - \ell\frac{v}{c}} \\
m' &= \frac{m}{\beta\left[1 - \ell\frac{v}{c}\right]} \\
n' &= \frac{n}{\beta\left[1 - \ell\frac{v}{c}\right]} \quad .
\end{aligned}\right\} \qquad (4)$$

We will interpret the formula for ω' in two different ways, depending on whether we consider the observer as moving and the (infinitely distant) source of light at rest, or vice versa.

[31] 1. If an observer is moving with velocity v relative to an infinitely distant source of light of frequency ν in such a way that the connecting line "source of light - observer" forms an angle φ with the observer's velocity as referred to a coordinate system at rest relative to the source of light, then the frequency ν' of the source of light perceived by the observer is given by the equation

$$\nu' = \nu \frac{1 - \cos \varphi\frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}} \; .$$

2. If a source of light of frequency ν_0 relative to a co-moving system moves such that the connecting line "source of light - observer" forms an angle φ with the velocity of the light source as referred to a system at rest relative to the observer, then the frequency ν of the source of light perceived by the observer is given by the equation

$$\nu = \nu_0 \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 - \cos\varphi \frac{v}{c}} . \qquad (4a)$$

The last two equations express Doppler's principle in its general form; the last equation shows how the observable frequency of the light emitted (or absorbed) by canal rays depends on the velocity of motion of the ions that form the rays and on the direction of sighting.

If the angles between the wave normal (ray direction) and the direction [32] of relative motion of S' with respect to S (i.e., with the x- and x'-axis, respectively) are called φ and φ', respectively, the equation for ℓ' takes the form

$$\cos \varphi' = \frac{\cos \varphi - \frac{v}{c}}{1 - \cos \varphi \frac{v}{c}} .$$

This equation shows the effect of the relative motion of the observer on the apparent location of an infinitely distant source of light (aberration).

In addition we will also examine how fast light is propagated in a medium that is moving in the direction of the light ray. Let the medium be at rest relative to the system S', and let the light vector be proportional to

$$\sin \omega' \left[t' - \frac{x'}{V'} \right]$$

or to

$$\sin \omega \left[t - \frac{x}{V} \right] ,$$

respectively, depending on whether the process is referred to S' or S.

The transformation equations yield

$$\omega = \beta\omega' \left[1 + \frac{v}{V'} \right]$$

$$\frac{\omega}{V} = \beta\frac{\omega'}{V'} \left[1 + \frac{V' v}{c^2} \right] .$$

Here V' should be viewed as a function of ω' known from the optics of stationary bodies. Dividing these equations, one obtains

$$V = \frac{V' + v}{1 + \frac{V'v}{c^2}} .$$

This equation could also have been obtained directly by applying the addition theorem for velocities.[1] If V' is to be considered as known, the last equation solves the problem completely. However, if only the frequency (ω) referred to the "stationary" system S is to be considered as known, as for example in the well-known experiment by Fizeau, then the two foregoing equations have to be used in conjunction with the relation between ω' and V' in order to determine the three unknowns ω', V', and V.

Further, if G or G' is the group velocity referred to S or S', respectively, then, according to the addition theorem for velocities,

$$G = \frac{G' + v}{1 + \frac{G'v}{c^2}} .$$

Since the relation between G' and ω' can be obtained from the optics of stationary bodies,[2] and ω' can be calculated from ω according to the foregoing, the group velocity G can be calculated even if only the frequency of light relative to S and the body's velocity of motion are given.

---

[33] [1] Cf. M. Laue, *Ann. d. Phys.* 23 (1907): 989.

[34] [2] Because $G' = \dfrac{V'}{1 + \frac{1}{V'}\frac{dV'}{d\omega'}}$.

II. ELECTRODYNAMIC PART

§7. *Transformation of the Maxwell-Lorentz equations* [35]

We start from the equations

$$\left.\begin{aligned} \frac{1}{c}\left\{u_x\rho + \frac{\partial X}{\partial t}\right\} &= \frac{\partial N}{\partial y} - \frac{\partial M}{\partial z} \\ \frac{1}{c}\left\{u_y\rho + \frac{\partial Y}{\partial t}\right\} &= \frac{\partial L}{\partial z} - \frac{\partial N}{\partial x} \\ \frac{1}{c}\left\{u_z\rho + \frac{\partial Z}{\partial t}\right\} &= \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y} \end{aligned}\right\} \qquad (5)$$

$$\left.\begin{aligned} \frac{1}{c}\frac{\partial L}{\partial t} &= \frac{\partial Y}{\partial z} - \frac{\partial Z}{\partial y} \\ \frac{1}{c}\frac{\partial M}{\partial t} &= \frac{\partial Z}{\partial x} - \frac{\partial X}{\partial z} \\ \frac{1}{c}\frac{\partial N}{\partial t} &= \frac{\partial X}{\partial y} - \frac{\partial Y}{\partial x} \end{aligned}\right\} \qquad (6)$$

In these equations,

(X,Y,Z) is the vector of electric field strength,
(L,M,N) is the vector of magnetic field strength,
$\rho = \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}$ is the 4π-fold electric density,
(u_x,u_y,u_z) is the velocity vector of electricity.

These equations, together with the assumption that the electrical masses are unchangeably attached to small rigid bodies (ions, electrons), form the basis of Lorentz's electrodynamics and optics of moving bodies.

If these equations, which shall hold with respect to the system S, are transformed by means of the transformation equations (1) to the moving system S', which is moving relative to S as in the previous considerations, then the following equations are obtained:

$$\left.\begin{aligned} \frac{1}{c}\left\{u'_x\rho' + \frac{\partial X'}{\partial t'}\right\} &= \frac{\partial N'}{\partial y'} - \frac{\partial M'}{\partial z'} \\ \frac{1}{c}\left\{u'_y\rho' + \frac{\partial Y'}{\partial t'}\right\} &= \frac{\partial L'}{\partial z'} - \frac{\partial N'}{\partial x'} \\ \frac{1}{c}\left\{u'_z\rho' + \frac{\partial Z'}{\partial t'}\right\} &= \frac{\partial M'}{\partial x'} - \frac{\partial L'}{\partial z'} \end{aligned}\right\} \tag{5'}$$

$$\left.\begin{aligned} \frac{1}{c}\frac{\partial L'}{\partial t'} &= \frac{\partial Y'}{\partial z'} - \frac{\partial Z'}{\partial y'} \\ \frac{1}{c}\frac{\partial M'}{\partial t'} &= \frac{\partial Z'}{\partial x'} - \frac{\partial X'}{\partial z'} \\ \frac{1}{c}\frac{\partial N'}{\partial t'} &= \frac{\partial X'}{\partial y'} - \frac{\partial Y'}{\partial x'} \end{aligned}\right\} \tag{6'}$$

where we have put

$$\left.\begin{aligned} X' &= X \\ Y' &= \beta\left[Y - \frac{v}{c}N\right] \\ Z' &= \beta\left[Z + \frac{v}{c}M\right] \end{aligned}\right\} \tag{7a}$$

$$\left.\begin{aligned} L' &= L \\ M' &= \beta\left[M + \frac{v}{c}Z\right] \\ N' &= \beta\left[N - \frac{v}{c}Y\right] \end{aligned}\right\} \tag{7b}$$

$$\rho' = \frac{\partial X'}{\partial x'} + \frac{\partial Y'}{\partial y'} + \frac{\partial Z'}{\partial z'} = \beta\left[1 - \frac{vu_x}{c^2}\right]\rho \tag{8}$$

$$\left.\begin{aligned} u'_x &= \frac{u_x - v}{1 - \frac{u_x v}{c^2}} \\ u'_y &= \frac{u_y}{\beta\left[1 - \frac{u_x v}{c^2}\right]} \\ u'_z &= \frac{u_x}{\beta\left[1 - \frac{u_x v}{c^2}\right]} \,. \end{aligned}\right\} \qquad (9)$$

The equations obtained have the same form as equations (5) and (6). On the other hand, it follows from the principle of relativity that the electrodynamic processes obey the same laws when they are related to S' as when they are related to S. From this we conclude that X', Y', Z' and L', M', N', respectively, are nothing else but the components of the S'-related electric and magnetic field strength.[1] Furthermore, inversion of equations (3) shows that the quantities u'_x, u'_y, u'_z in equations (9) equal the S'-related velocity components of the electricity, and hence ρ' is the S'-related density of electricity. Thus the electrodynamic basis of the Maxwell-Lorentz theory agrees with the principle of relativity.

Regarding the interpretation of equations (7a) we note the following. [37] Imagine a pointlike quantity of electricity that is at rest relative to S and is of magnitude "one" with respect to S, i.e., exerts a force of 1 dyne on an equal quantity of electricity located at a distance of 1 cm and at rest with respect to S. According to the principle of relativity, this electric mass also equals "one" when it is at rest relative to S' and is examined from S'.[2] If this quantity of electricity is at rest relative to S, then (X,Y,Z) is by definition equal to the force acting upon it, which could be measured, for example, by a spring balance at rest relative to S. The vector (X',Y',Z') has the analogous meaning with regard to S'.

---

[1] Though the agreement between the equations found and equations (5) and (6) leaves open the possibility that the quantities X', etc., differ by a constant factor from the S'-referred field strengths, it is easy to show by a method very similar to that employed in §3 for the function $\varphi(v)$ that this factor must equal 1. [36]

[2] This conclusion rests further on the assumption that the magnitude of an electric mass does not depend on the prehistory of its motion.

According to equations (7a) and (7b), electric or magnetic field strengths do not have an existence per se, since it may depend on the choice of the coordinate system whether an electric or magnetic field strength is or is not present at a location (more exactly: spatial-temporal environment of a point event). Further, if one introduces a reference system that is at rest with respect to the electric mass, one sees that the "electromotive" forces introduced hitherto that act upon an electric mass moving in a magnetic field are nothing else but "electric" forces. This makes the questions as to the seat of those electromotive forces (in unipolar machines) pointless, since the answer varies depending on the choice of the state of motion of the reference system used.

[38] The meaning of equation (8) can be seen from the following: Let an electrically charged body be at rest relative to S'. Its total charge ϵ' with respect to S' is then $\int \frac{\rho'}{4\pi}\, dx'dy'dz'$. How large is its total charge ϵ at some time t of S?

It follows from the last three of equations (1) that the following relation holds for constant t:

$$dx'dy'dz' = \beta dxdydz \ .$$

In our case equation (8) reads

$$\rho' = \frac{1}{\beta}\rho .$$

From these two equations it follows that we must have

$$\epsilon' = \epsilon \ .$$

Equation (8) thus states that the electric mass is a quantity that is independent of the state of motion of the reference system. Thus, if the charge of some body in motion is constant from the standpoint of a reference system moving with it, then it is also constant with respect to any other reference system.

With the help of equations (1), (7), (8), and (9), all problems involving the electrodynamics and optics of moving bodies in which only velocities, but not accelerations, play an essential role can be reduced to a series of problems involving the electrodynamics or optics of stationary bodies.

We shall illustrate the application of the equations developed here by one additional simple example. Let a plane light wave traveling through vacuum be described by the following equations with respect to S:

$$\begin{aligned} X &= X_0 \sin \Phi & \quad L &= L_0 \sin \Phi \\ Y &= Y_0 \sin \Phi & \quad M &= M_0 \sin \Phi & \quad \Phi &= \omega\left[t - \frac{\ell x + my + nz}{c}\right] \\ Z &= Z_0 \sin \Phi & \quad Z &= N_0 \sin \Phi \end{aligned}$$ [39]

We ask about the constitution of this wave when referred to the system S'.

Application of the transformation equations (1) and (7) yields

$$\begin{aligned} X' &= X_0 \sin \Phi' & \qquad L' &= L_0 \sin \Phi' \\ Y' &= \beta\left[Y_0 - \frac{v}{c} N_0\right]\sin \Phi' & \qquad M' &= \beta\left[M_0 + \frac{v}{c} Z_0\right]\sin \Phi' \\ Z' &= \beta\left[Z_0 + \frac{v}{c} M_0\right]\sin \Phi' & \qquad N' &= \beta\left[N_0 - \frac{v}{c} Y_0\right]\sin \Phi' \end{aligned}$$

$$\Phi' = \omega'\left[t' - \frac{\ell' x' + m' y' + n' z'}{c}\right] .$$ [40]

From the requirement that the functions X', etc., must satisfy equations (5') and (6'), it follows that the wave normal, electric force, and magnetic force are mutually perpendicular with respect to S' as well, and that the latter two are equal. The relations following from the identity $\Phi = \Phi'$ were discussed in §6; only the amplitude and state of polarization of the wave relative to S' remain to be determined here.

We choose the X-Y-plane parallel to the wave normal and deal first with the case in which the electric vibration is parallel to the Z-axis. In that case we have to set

$$\begin{aligned} X_0 &= 0 & L_0 &= -A \sin \varphi \\ Y_0 &= 0 & M_0 &= -A \cos \varphi \\ Z_0 &= A & N_0 &= 0, \end{aligned}$$

where φ denotes the angle between the wave normal and the X-axis. From the above it follows that

$$\begin{aligned} X' &= 0 & L' &= -A \sin \varphi \sin \Phi' \\ Y' &= 0 & M' &= \beta\left[-\cos \varphi + \frac{v}{c}\right] A \sin \Phi' \\ Z' &= \beta\left[1 - \frac{v}{c} \cos \varphi\right] A \sin \Phi' & N' &= 0\ . \end{aligned}$$

Thus, if A' denotes the amplitude of the wave with respect to S', we will have

$$A' = A \frac{1 - \frac{v}{c} \cos \varphi}{\sqrt{1 - \frac{v^2}{c^2}}}\ . \tag{10}$$

The same relation obviously holds for the special case that the *magnetic* force is perpendicular to the direction of relative motion and the wave normal. Since the general case can be constructed from these two special cases by superposition, it follows that relation (10) is valid in general if a new reference system S' is introduced, and that the angle between the plane of polarization and a plane parallel to the wave normal and to the direction of the relative motion is the same in the two reference systems.

III. MECHANICS OF THE MATERIAL POINT (ELECTRON)

[41]

§8. *Derivation of the equations of motion of the (slowly accelerated) material point, or electron*

Let a particle endowed with an electric charge ϵ (which we shall call "electron" in the following) move in an electromagnetic field, and let us assume the following about the law of motion of this particle:

If at a given point of time the electron is at rest relative to a (nonaccelerated) reference system S', its motion relative to S' will proceed according to the following equations in the next instant of time:

$$\mu \frac{d^2 x'_0}{dt'^2} = \epsilon X'$$

$$\mu \frac{d^2 y'_0}{dt'^2} = \epsilon Y'$$

$$\mu \frac{d^2 z'_0}{dt'^2} = \epsilon Z' \quad ,$$

where x'_0, y'_0, z'_0 denote the coordinates of the electron with respect to S', and μ is a constant which we call the mass of the electron.

We introduce a system S, relative to which S' is in motion as in our preceding analyses, and transform our equations of motion using the transformation equations (1) and (7a).

In our case, the former will read

$$\begin{aligned} t' &= \beta\left[t - \frac{v}{c^2} x_0\right] \\ x'_0 &= \beta(x_0 - vt) \\ y'_0 &= y_0 \\ z'_0 &= z_0 \ . \end{aligned}$$

Setting $\frac{dx_0}{dt} = \dot{x}_0$, etc., we obtain from these equations

$$\frac{dx'_0}{dt'} = \frac{\beta(\dot{x}_0 - v)}{\beta\left[1 - \frac{v\dot{x}_0}{c^2}\right]} \text{ , etc.,}$$

$$\frac{d^2 x'_0}{dt'^2} = \frac{\frac{d}{dt}\left\{\frac{dx'_0}{dt'}\right\}}{\beta\left[1 - \frac{v\dot{x}_0}{c^2}\right]} = \frac{1}{\beta} \frac{\left[1 - \frac{v\dot{x}_0}{c^2}\right]\ddot{x}_0 + (\dot{x}_0 - v)\frac{v\ddot{x}_0}{c^2}}{\left[1 - \frac{v\dot{x}_0}{c^2}\right]} \text{ , etc.}$$ [42]

Inserting these expressions into the above equations after having put $\dot{x}_0 = v$, $\dot{y}_0 = 0$, $\dot{z}_0 = 0$, while at the same time substituting X', Y', Z' by means if equations (7a), one gets

$$\begin{aligned} \mu\beta^3\ddot{x}_0 &= \epsilon X \\ \mu\beta\ddot{y}_0 &= \epsilon\left[Y - \frac{v}{c}N\right] \\ \mu\beta\ddot{z}_0 &= \epsilon\left[Z + \frac{v}{c}M\right] . \end{aligned}$$

These equations are the equations of motion of the electron for the case when $\dot{x}_0 = v$, $\dot{y}_0 = 0$, $\dot{z}_0 = 0$ at the instant in question. On the left-hand sides, then, v may be replaced by the velocity q defined by the equation

$$q = \sqrt{\dot{x}_0^2 + \dot{y}_0^2 + \dot{z}_0^2}$$

and on the right-hand sides v may be replaced by $\dot{x}_0$. Further, we add in the appropriate places the terms obtained from $\frac{\dot{x}_0}{c}M$ and $-\frac{\dot{x}_0}{c}N$ by cyclic permutation, which vanish in the particular case under consideration. Omitting the subscript in x_0, etc., we obtain the following equations, which in the particular case under consideration are equivalent to the equations
[43] given above:

$$\left.\begin{aligned} \frac{d}{dt}\left\{\frac{\mu\dot{x}}{\sqrt{1 - \frac{q^2}{c^2}}}\right\} &= K_x \\ \frac{d}{dt}\left\{\frac{\mu\dot{y}}{\sqrt{1 - \frac{q^2}{c^2}}}\right\} &= K_y \\ \frac{d}{dt}\left\{\frac{\mu\dot{z}}{\sqrt{1 - \frac{q^2}{c^2}}}\right\} &= K_z , \end{aligned}\right\} \qquad (11)$$

where we have put

$$\left.\begin{aligned}K_x &= \epsilon\left\{X + \frac{\dot{y}}{c}N - \frac{\dot{z}}{c}M\right\}\\ K_y &= \epsilon\left\{Y + \frac{\dot{z}}{c}L - \frac{\dot{x}}{c}N\right\}\\ K_z &= \epsilon\left\{Z + \frac{\dot{x}}{c}M - \frac{\dot{y}}{c}L\right\}\ .\end{aligned}\right\} \qquad (12)$$

These equations do not change their form with the introduction of a new coordinate system with differently directed axes, which is relatively at rest. Hence they are valid in general and not only when $\dot{y} = \dot{z} = 0$.

The vector (K_x, K_y, K_z) shall be called the force acting on the material point. If q^2 is vanishingly small compared with c^2, then according to equations (11) K_x, K_y, K_z reduce to the force components according to Newton's definition. In the next section it will be shown that in other respects, too, that vector plays the same role in relativity mechanics as the force does in classical mechanics.

We shall maintain equations (11) also in the case that the force exerted on the mass point is not of electromagnetic nature. In the latter case equations (11) do not have a physical content but are rather to be understood as defining equations of the force.

§9. *Motion of the mass point and the principles of mechanics*

If equations (5) and (6) are successively multiplied by $\frac{X}{4\pi}, \frac{Y}{4\pi} \cdots \frac{N}{4\pi}$, and integrated over a space on whose boundaries the field strengths vanish, one obtains

$$\int \frac{\rho}{4\pi}(u_x X + u_y Y + u_z Z)\,d\omega + \frac{dE_e}{dt} = 0, \qquad (13)$$

where

$$E_e = \int \left[\frac{1}{8\pi}(X^2 + Y^2 + Z^2) + \frac{1}{8\pi}(L^2 + M^2 + N^2)\right] d\omega$$

is the electromagnetic energy of the space under consideration. According to the energy principle, the first term of equation (13) equals the energy delivered by the electromagnetic field to the carrier of the electric masses per unit time. If electric masses are rigidly bound to a material point (electron), then their part in the above term equals the expression

$$\epsilon(X\dot{x} + Y\dot{y} + Z\dot{z}) \ ,$$

where (X,Y,Z) denotes the *external* electric field strength, i.e., the field strength minus that part which is due to the charge of the electron itself. Using equations (12), this expression becomes

$$K_x\dot{x} + K_y\dot{y} + K_z\dot{z} \ .$$

Thus the vector (K_x,K_y,K_z) denoted as "force" in the last paragraph has the same relation to the work performed as in Newtonian mechanics.

[44] Thus, if one successively multiplies equations (11) by $\dot{x}$, $\dot{y}$, $\dot{z}$, then adds and integrates over time, this must yield the kinetic energy of the material point (electron). One obtains

$$\int (K_x\dot{x} + K_y\dot{y} + K_z\dot{z})dt = \frac{\mu c^2}{\sqrt{1 - \frac{q^2}{c^2}}} + \text{const.} \qquad (14)$$

By this we have demonstrated that the equations of motion (11) are in accord with the energy principle. We will now show that they are also in accord with the principle of conservation of momentum.

Successively multiplying the second and third of equations (5) and the second and third of equations (6) by $\frac{N}{4\pi}$, $\frac{-M}{4\pi}$, $\frac{-Z}{4\pi}$, $\frac{Y}{4\pi}$, adding them and integrating over a space at whose boundaries the field strengths vanish, we obtain

$$\frac{d}{dt}\left[\int \frac{1}{4\pi c}(YN - ZM)\,d\omega\right] + \int \frac{\rho}{4\pi}\left[X + \frac{u_y}{c} N - \frac{u_z}{c} M\right] d\omega = 0 \qquad (15)$$

or, according to equations (12),

$$\frac{d}{dt}\left[\iint \frac{1}{4\pi c}(YN - ZM)\,d\omega\right] + \Sigma K_x = 0 \; . \tag{15a}$$

If the electric masses are bound to freely moving material points (electrons), this equation becomes by virtue of (11)

$$\frac{d}{dt}\left[\iint \frac{1}{4\pi c}(YN - ZM)\,d\omega + \Sigma\right] \frac{\mu\dot{x}}{\sqrt{1 - \frac{q^2}{c^2}}} = 0 \; . \tag{15b}$$ [45]

In combination with the equations obtained by cyclic permutation, this equation expresses the principle of conservation of momentum in the case considered here. Thus the quantity $\xi = \frac{\mu\dot{x}}{\sqrt{1 - \frac{v^2}{c^2}}}$ plays the role of the [46] momentum of the material point, and in accordance with equations (11) we have

$$\frac{d\xi}{dt} = K_x$$

as in classical mechanics. The possibility of introducing a momentum of the material point is based on the fact that in the equations of motion the force, i.e., the second term of equation (15), can be represented as a time derivative.

Further, one sees immediately that our equations of motion of the material point can be given the form of Lagrange's equations of motion; for, according to equations (11), we have

$$\frac{d}{dt}\left[\frac{\partial H}{\partial \dot{x}}\right] = K_x \; , \text{ etc.}$$

where we have put

$$H = -\mu c^2 \sqrt{1 - \frac{q^2}{c^2}} + \text{const.}$$ [47]

The equations of motion can also be represented in the form of Hamilton's principle

$$\int_{t_0}^{t_1} (dH + A)\,dt = 0 \; ,$$

where the time t and the initial and final position remain unvaried, and where A denotes the virtual work

$$A = K_x \partial x + K_y \partial y + K_z \partial z \ .$$

Finally, we establish Hamilton's canonical equations of motion. This is done by introducing the "momentum coordinates" (components of the momentum) ξ, η, ζ, setting as above

[48] $$\xi = \frac{\partial H}{\partial \dot{x}} = \frac{\mu x}{\sqrt{1 - \frac{q^2}{c^2}}} \ , \text{ etc.}$$

If one considers the kinetic energy L as a function of ξ, η, ζ, and sets $\xi^2 + \eta^2 + \zeta^2 = \rho^2$, one obtains

$$L = \mu c^2 \sqrt{1 + \frac{\rho^2}{\mu^2 c^2}} + \text{const.} \ ,$$

and Hamilton's equations of motion become

$$\frac{d\xi}{dt} = K_x \qquad \frac{d\eta}{dt} = K_y \qquad \frac{d\zeta}{dt} = K_z$$

$$\frac{dx}{dt} = \frac{\partial L}{\partial \xi} \qquad \frac{dy}{dt} = \frac{\partial L}{\partial \eta} \qquad \frac{dz}{dt} = \frac{\partial L}{\partial \zeta} \ .$$

[49] *§10. On the possibility of an experimental test of the theory of motion of the material point. Kaufmann's investigation*

A prospect of comparison with experience of the results derived in the last section exists only where the moving electrically charged mass points possess velocities whose square is not negligible compared to c^2. This condition is satisfied in the cases of the faster cathode rays and the electron rays (β-rays) emitted by radioactive substances.

There are three quantities for electron rays whose mutual relationships can be the subject of a more detailed experimental investigation, i.e., the generating potential or the kinetic energy of the rays, the deflectability by an electric field, and the deflectability by a magnetic field.

According to (14), the generating potential Π is given by the formula

$$\Pi\epsilon = \mu\left\{\frac{c^2}{\sqrt{1 - \frac{q^2}{c^2}}} - 1\right\} .$$ [50]

To calculate the two other quantities, we use the last of equations (11) for the case when the motion is momentarily parallel to the X-axis; denoting the absolute value of the electron's charge by ϵ, one obtains

$$-\frac{d^2z}{dt^2} = \frac{\epsilon}{\mu}\sqrt{1 - \frac{q^2}{c^2}}\left[Z + \frac{q}{c}M\right] .$$

If Z and M are the only deflecting field components, and hence the bending takes place in the XZ-plane, the radius of curvature R of the trajectory is given by $\frac{q^2}{R} = \left[\frac{d^2z}{dt^2}\right]$. Hence, if the electric and magnetic deflectability are defined as the quantities $A_e = \frac{1}{R} : Z$ and $A_m = \frac{1}{R} : M$ respectively for the case that *only* one deflecting electric or *only* one magnetic field component is present, one has

$$A_e = \frac{\epsilon}{\mu}\frac{\sqrt{1 - \frac{q^2}{c^2}}}{q^2}$$

$$A_m = \frac{\epsilon}{\mu}\frac{\sqrt{1 - \frac{q^2}{c^2}}}{cq} .$$

In the case of cathode rays all three quantities Π, A_e, and A are possible candidates for measurement; however, no investigations with sufficiently fast cathode rays have yet been performed. In the case of β-rays, only the quantities A_e and A_m are (in practice) accessible to [51] observation. Mr. W. Kaufmann ascertained the relation between A_m and A_e for β-rays emitted by a radium bromide granule with admirable care.[1]

---

[1] W. Kaufmann, "Über die Konstitution des Elektrons" [On the constitution of the electron]. *Ann. d. Phys.* 19 (1906). Both figures are taken from Kaufmann's paper. [52]

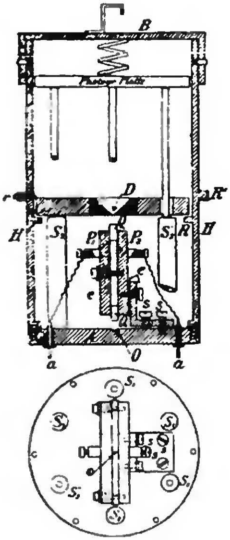


Fig. 1 (actual size)

[53] His apparatus, whose main parts are depicted in their actual size in Fig. 1, consisted essentially of a lightproof brass casing *H* placed within an evacuated glass vessel, with a radium granule placed in a small well *O* in the floor *A* of the casing. The β-rays emanating from the radium pass through the gap between two condenser plates P_1 and P_2, cross the diaphragm *D*, whose diameter is 0.2 mm, and then fall on the photographic plate. The rays were deflected both by an electric field formed between the condenser plates P_1 and P_2 and by a magnetic field of the same direction (produced by a large permanent magnet), perpendicular to that direction so that rays of the same velocity marked a point, and the aggregate of the particles of different velocities marked a curve on the plate.

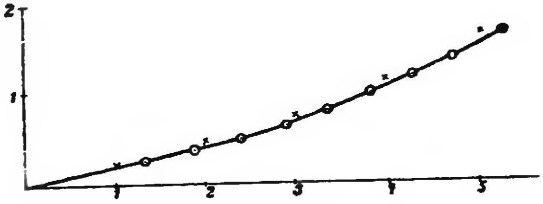


Fig. 2.

Fig. 2 shows this curve[1] which, up to the scale for the abscissa and ordinate, represents the relation between A_m (abscissa) and A_e (ordinate). The little crosses above the curve indicate the curve calculated according to the theory of relativity, if the value of $\frac{\epsilon}{\mu}$ is taken as $1.878 \cdot 10^7$. [54] [56]

In view of the difficulties involved in the experiment one would be inclined to consider the agreement as satisfactory. However, the deviations are systematic and considerably beyond the limits of error of Kaufmann's experiment. That the *calculations* of Mr. Kaufmann are error-free is shown by the fact that, using another method of calculation, Mr. Planck arrived at results that are in full agreement with those of Mr. Kaufmann.[2]

Only after a more diverse body of observations becomes available will it be possible to decide with confidence whether the systematic deviations are due to a not yet recognized source of errors or to the circumstance that the foundations of the theory of relativity do not correspond to the facts. [58]

---

[1] The units given in the graph denote millimeters on the photographic plate. The plotted curve is not exactly the one observed, but rather the curve "reduced to infinitesimally small deflections." [55]

[2] Cf. M. Planck, *Verhandl. d. Deutschen Phys. Ges.* VIII, no. 20 (1906); IX, no. 14 (1907). [57]

It should also be mentioned that Abraham's[1] and Bucherer's[2] theories of the motion of the electron yield curves that are significantly closer to the observed curve than the curve obtained from the theory of relativity. However, the probability that their theories are correct is rather small, in my opinion, because their basic assumptions concerning the dimensions of the moving electron are not suggested by theoretical systems that encompass larger complexes of phenomena.

IV. ON THE MECHANICS AND THERMODYNAMICS OF SYSTEMS

[61] §*11. On the dependence of mass upon energy*

We consider a physical system surrounded by a casing impenetrable to radiation. Suppose that the system floats freely in space and is not subjected to any other forces except the effects of electric and magnetic forces of the surrounding space. Through the latter, energy may be transferred to the system in the form of work and heat, and this energy may undergo conversions of some sort in the interior of the system. In accordance with (13), the energy absorbed by the system is given by the following expressions when referred to the system S:

$$\int dE = \int dt \int \frac{\rho}{4\pi}(X_a u_x + Y_a u_y + Z_a u_z)d\omega \text{ ,}$$

where (X_a, Y_a, Z_a) denotes the field vector of the external field (which is not included in the system) and $\frac{\rho}{4\pi}$ the electric density in the casing. We transform this expression by inverting equations (7a), (8), and (9), taking into account that according to equations (1) the functional determinant

$$\frac{D(x', y', z', t')}{D(x, y, z, t)}$$

---

[59] [1] M. Abraham, *Gött. Nachr.* 1902.

[60] [2] A. H. Bucherer, *Math. Einführung in die Elektronentheorie* [Mathematical introduction to the electron theory], Leipzig, 1904, p. 58.

equals one. We thus obtain

$$\int dE = \beta \iint \frac{\rho'}{4\pi}(u'_x X'_a + u'_y Y'_a + u'_z Z'_a)\,d\omega'\,dt' + \beta v \iint \frac{\rho'}{4\pi}\left[X'_a + \frac{u'_y}{c} N'_a - \frac{u'_z}{c} M'_a\right] d\omega'\,dt' ,$$

or, since the energy principle must hold with regard to S' as well, in easily comprehensible notation

$$dE = \beta dE' + \beta v \int [\Sigma K'_x]\,dt' \; . \tag{16}$$

We shall now apply this equation to the case that the system under consideration moves uniformly such that as a whole it is at rest relative to the reference system S'. Then, provided that the parts of the system move so slowly relative to S' that the squares of the velocities relative to S' are negligible compared with c^2, we can apply the principles of Newtonian mechanics with regard to S'. Thus, according to the center-of-mass theorem, the system under consideration (or, more accurately, its center of gravity) can remain at rest permanently only if for each t'

$$\Sigma K'_x = 0 \; .$$

Nevertheless, the second term on the right-hand side of equation (16) does not necessarily vanish, because the integration over time is to be performed between two specific values of t and not of t'.

However, if at the beginning and end of the time span considered no external forces act upon the system of bodies, that term vanishes and we obtain simply

$$dE = \beta \cdot dE' \; .$$

First of all, we conclude from this equation that the energy of a (uniformly) moving system not affected by external forces is a function of two variables, i.e., the energy E_0 of the system relative to a reference system moving with it,[1] and the translational velocity q of the system, and we obtain

$$\frac{\partial E}{\partial E_0} = \frac{1}{\sqrt{1 - \frac{q^2}{c^2}}} .$$

From this it follows that

[62] $$E = \frac{1}{\sqrt{1 - \frac{q^2}{c^2}}} E_0 + \varphi(q) ,$$

where $\varphi(q)$ is a function of q that is unknown for the time being. The case that E_0 equals 0, i.e., that the energy of the moving system is a function of the velocity q alone, has already been examined in §§ 8 and 9. From equation (14) it follows immediately that we have to put

$$\varphi(q) = \frac{\mu c^2}{\sqrt{1 - \frac{q^2}{c^2}}} + \text{const.}$$

Thus we obtain

$$E = \left[\mu + \frac{E_0}{c^2}\right] \frac{c^2}{\sqrt{1 - \frac{q^2}{c^2}}} , \tag{16a}$$

where the integration constant has been omitted. A comparison of this expression for E with the expression for the kinetic energy of the material point contained in equation (14) shows that the two expressions have the same form; with regard to the dependence of the energy on the translational velocity, the physical system under consideration behaves like a material point of mass M, where M depends on the system's energy content E_0 according to the formula

$$M = \mu + \frac{E_0}{c^2} . \tag{17}$$

This result is of extraordinary theoretical importance because the inertial mass and the energy of a physical system appear in it as things of

---

[1] Here, as well as in the following, we use a symbol with the subscript "0" to indicate that the quantity in question refers to a reference system that is at rest relative to the physical system considered. Since the system considered is at rest relative to S', we can replace E' by E_0 here.

the same kind. With respect to inertia, a mass μ is equivalent to an energy content of magnitude μc^2. Since we can arbitrarily assign the zero-point of E_0, we are not even able to distinguish between a system's "actual" and "apparent" mass without arbitrariness. It seems far more natural to consider any inertial mass as a reserve of energy.

According to our result, the law of the constancy of mass applies to a single physical system only when its energy remains constant; it is then equivalent to the energy principle. To be sure, the changes experienced by the mass of physical systems during the familiar physical processes are always immeasurably small. For example, the decrease in mass of a system that gives off 1000 gram-calories amounts to 4.6×10^{-11} gram.

The radioactive decay of a substance is accompanied by the release of enormous amounts of energy; is the reduction of mass in such a process not large enough to be detectable?

Mr. Planck writes about this: "According to J. Precht[1] 1 gram-atom of [63]
radium, if surrounded by a sufficiently thick layer of lead, releases $134.4 \times 225 = 30{,}240$ gram-calories per hour. According to (17) this amounts
to a decrease in mass of [65]

$$\frac{30240 \cdot 419 \cdot 10^5}{9 \cdot 10^{20}} \text{ gr} = 1.41 \times 10^{-6} \text{ mg}$$

per hour or 0.012 mg per year. Of course, this amount is still so tiny, especially in view of the high atomic weight of radium, that it may well be outside the experimentally accessible range for the time being." The obvious question arises whether it would not be possible to reach one's goal by using an indirect method. If M is the atomic weight of the disintegrating atom, and m_1, m_2, etc., are the atomic weights of the end products of radioactive disintegration, then we must have

$$M - \Sigma m = \frac{E}{c^2} \, ,$$

where E denotes the energy produced during the disintegration of one gram-atom; this can be calculated if the energy developed per unit time during

---

[1] J. Precht, *Ann. d. Phys.* 21 (1906): 599. [64]

stationary disintegration and the average disintegration constant of the atom are known. Whether the method can be successfully applied depends primarily on whether there are radioactive reactions for which $\frac{M - \Sigma m}{M}$ is not too small compared to 1. In the above-mentioned case of radium one obtains--if its lifetime is taken to be 2600 years--approximately

$$\frac{M - \Sigma m}{M} = \frac{12 \cdot 10^{-6} \cdot 2600}{250} = 0.00012 \ .$$

Thus, if the lifetime of radium has been ascertained with fair accuracy, we could check our relations if we knew the atomic weights involved with an accuracy of five places. This, of course, is impossible. However, it is possible that radioactive processes will be detected in which a significantly higher percentage of the mass of the original atom will be converted into the energy of a variety of radiations than in the case of radium. At least it seems reasonable to imagine that the energy produced during the disintegration of an atom varies at least as much from substance to substance as does the [66] rate of disintegration.

It has been tacitly assumed above that such a change in mass can be measured by the instrument we usually use for measuring masses, i.e., by the balance, and hence that the relationship

$$M = \mu + \frac{E_0}{c^2}$$

holds not only for the inertial mass but also for the gravitational mass, or, in other words, that a system's inertia and weight are strictly proportional [67] under all circumstances. We would also have to assume, for example, that radiation enclosed in a cavity possesses not only inertia but also weight. But this proportionality between the inertial and gravitational mass holds [68] without exception for all bodies with the accuracy obtained so far, so that we must assume its general validity until it is proven otherwise. We are going to find a new argument in support of this assumption in the last section of this paper.

§12. *Energy and momentum of a moving system*

As in the previous section, we once again consider a system that floats freely in space and is enveloped by a casing impervious to radiation. We again denote the field strengths of the external electromagnetic field which mediates the energy exchange with other systems by X_a, Y_a, Z_a, etc. We can apply to this external field the reasoning that led us to formula (15), so that we obtain

$$\frac{d}{dt}\left[\int \frac{1}{4\pi c}\,(Y_a N_a - Z_a M_a)\,d\omega\right] + \int \frac{\rho}{4\pi}\left[X_a + \frac{u_y}{c}N_a - \frac{u_z}{c}M_a\right]d\omega = 0\ .$$

We shall now assume that the principle of conservation of momentum is universally valid. In that case it must be possible to represent the part of the second term of this equation that extends over the casing of the system as the time derivative of a quantity G_x, which is completely determined by the instantaneous state of the system and which we denote as the X-component of the momentum of the system. We wish now to find the transformation law of the quantity G_x. Applying the transformation equations (1), (7), (8), and (9), we obtain in exactly the same way as in the previous section the relation

$$\int dG_x = \beta \int\int \frac{\rho'}{4\pi}\left[X'_a + \frac{u'_y}{c}N'_a - \frac{u'_z}{c}M'_a\right]d\omega'\cdot dt'$$

$$+ \frac{\beta v}{c^2}\int\int \frac{\rho'}{4\pi}(X'_a u'_x + Y'_a u'_y + Z'_a u'_z)\,d\omega\cdot dt'$$ [69]

or

$$dG_x = \beta\,\frac{v}{c^2}\,dE' + \beta\int \{\Sigma K'_x\}dt'\ . \qquad (18)$$

Again, let the body move without acceleration, so that it is permanently at rest with respect to S'. Then we have again

$$\Sigma K'_x = 0\ .$$

Although the limits of the time integral depend on x', the second term on the right-hand side of the equation vanishes again if the body is not exposed to external forces before and after the change under consideration. We then have

$$dG_x = \beta \frac{v}{c^2} dE' .$$

From this it follows that the momentum of a system not exposed to external forces is a function of only two variables, namely, the energy E_0 of the system relative to a reference system moving along with it, and the translation velocity q of the latter. We have

$$\frac{\partial G}{\partial E_0} = \frac{\frac{q}{c^2}}{\sqrt{1 - \frac{q^2}{c^2}}} .$$

This implies

[70] $$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} \cdot \left[\frac{E_0}{c^2} + \psi(q)\right] ,$$

where $\psi(q)$ is an as yet unknown function of q. Since $\psi(q)$ is in fact the momentum if the latter is determined by the velocity alone, we conclude from formula (15b) that

$$\psi(q) = \frac{\mu q}{\sqrt{1 - \frac{q^2}{c^2}}} .$$

We thus obtain

$$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} \left\{\mu + \frac{E_0}{c^2}\right\} . \tag{18a}$$

The only difference between this expression and the expression for the momentum of the material point is that μ has been replaced by $\left[\mu + \frac{E_0}{c^2}\right]$ in accordance with the result of the previous section.

[71] We will now determine the energy and momentum of a body at rest relative to S if the body is subjected to permanent external forces. Even though in that case we again have

$$\Sigma K'_x = 0 ,$$

for each t', the integral

$$\int [\Sigma K'_x]\, dt'$$

that appears in equations (16) and (18) does not vanish, because it has to be extended over two definite values of t rather than of t'. Since inversion of the first of equations (1) yields

$$t = \beta\left[t' + \frac{v}{c^2}\, x'\right] ,$$

the limits of integration over t' are given by

$$\frac{t_1}{\beta} - \frac{v}{c^2}\, x' \qquad \text{and} \qquad \frac{t_2}{\beta} - \frac{v}{c^2}\, x' ,$$

where t_1 and t_2 are independent of x', y', z'. Hence the limits of the time integration with respect to S' depend on the position of the points at which the forces are applied. We split the above integral into three integrals:

$$\int [\Sigma K'_x]\, dt' = \int\limits_{\frac{t_1}{\beta} - \frac{vx'}{c^2}}^{\frac{t_1}{\beta}} + \int\limits_{\frac{t_1}{\beta}}^{\frac{t_2}{\beta}} + \int\limits_{\frac{t_2}{\beta}}^{\frac{t_2}{\beta} - \frac{vx'}{c^2}} .$$

The second of these integrals vanishes because it has constant time limits. If in addition the forces K'_x can change arbitrarily fast, the other two integrals cannot be evaluated; then we cannot talk at all about the energy or momentum of the system while applying the principles used here.[1] However, if those forces change very little during time intervals of the order of magnitude $\frac{vx'}{c^2}$, we can put

---

[1] Cf. A. Einstein, *Ann. d. Phys.* 23 (1907), §2. [72]

$$\int_{\frac{t_1}{\beta}-\frac{vx'}{c^2}}^{\frac{t_1}{\beta}} (\Sigma K'_x)\,dt' = \Sigma K'_x \int_{\frac{t_1}{\beta}-\frac{vx'}{c^2}}^{\frac{t_1}{\beta}} dt' = \frac{v}{c^2}\,\Sigma x' K'_x \;.$$

After similar evaluation of the third integral one obtains

$$\int (\Sigma K'_x)\,dt' = -\,d\left\{\frac{v}{c^2}\,\Sigma x' K'_x\right\} .$$

Now the energy and momentum can be calculated from equations (16) and (18) without difficulty. One obtains

$$E = \left[\mu + \frac{E_0}{c^2}\right] \frac{c^2}{\sqrt{1-\frac{q^2}{c^2}}} - \frac{\frac{q^2}{c^2}}{\sqrt{1-\frac{q^2}{c^2}}}\,\Sigma(\delta_0 K_{0\delta}) \qquad (16b)$$

[73]

$$q = \frac{q}{\sqrt{1-\frac{q^2}{c^2}}} \left[\mu + \frac{E_0 - \Sigma(\delta_0 K_{0\delta})}{c^2}\right] , \qquad (18b)$$

where $K_{0\delta}$ denotes the component in the direction of motion of a force evaluated in a co-moving reference system and δ_0 denotes the distance, measured in the same system, between the point of application of that force and a plane perpendicular to the direction of motion.

If, as we shall assume henceforth, the external force consists of a pressure p_0 which is independent of the direction and always acts perpendicularly to the surface of the system, we will have in particular

$$\Sigma(\delta_0 K_{0\delta}) = -p_0 V_0 \;, \qquad (19)$$

where V_0 is the volume of a system evaluated in the co-moving reference [74] system. Equations (16b) and (18b) then take the form

$$E = \left[\mu + \frac{E_0}{c^2}\right] \frac{c^2}{\sqrt{1 - \frac{q^2}{c^2}}} + \frac{\frac{q^2}{c^2}}{\sqrt{1 - \frac{q^2}{c^2}}} p_0 V_0 \tag{16c}$$

$$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} \left[\mu + \frac{E_0 + p_0 V_0}{c^2}\right] . \tag{18c}$$

§13. *The volume and pressure of a moving system. Equations of motion* [75]

To determine the state of the system under consideration, we have used the quantities E_0, p_0, V_0, which are defined with respect to a reference system that moves along with the physical system. However, instead of these quantities we can also use the corresponding quantities that are defined with respect to the same reference system as the momentum G. To do this, we must examine how the volume and pressure change with the introduction of a new reference system.

Let a body be at rest with respect to the reference system S'. Let V' be its volume with respect to S', and V its volume with respect to S. It follows immediately from equations (2) that

$$\int dx \cdot dy \cdot dz = \sqrt{1 - \frac{v^2}{c^2}} \int dx' \cdot dy' \cdot dz'$$

or

$$V = \sqrt{1 - \frac{v^2}{c^2}} \cdot V' .$$

If we replace V' by V_0 and v by q in accordance with the notation used, we will get

$$V = \sqrt{1 - \frac{q^2}{c^2}} \cdot V_0 . \tag{20}$$

To obtain the transformation equations for the pressure forces, we must start from the transformation equations that apply to forces in general. Since we defined the moving forces in §8 in such a way that they can be replaced by the force effects of electromagnetic fields on electric masses, we can now restrict ourselves to determining the transformation equations of the latter.[1]

Let the quantity of electricity ϵ be at rest with respect to S'. In accordance with equations (12), the force acting upon it is then given by the equations

$$K_x = \epsilon X \qquad\qquad K'_x = \epsilon X'$$

$$K_y = \epsilon\left[Y - \frac{v}{c}N\right] \qquad\qquad K'_y = \epsilon Y'$$

$$K_z = \epsilon\left[Z + \frac{v}{c}M\right] \qquad\qquad K'_z = \epsilon Z' \ .$$

From these equations and equations (7a) it follows that

$$\left.\begin{aligned} K'_x &= K_x \\ K'_y &= \beta\cdot K_y \\ K'_z &= \beta\cdot K_z \ . \end{aligned}\right\} \tag{21}$$

These equations allow the calculation of forces if they are known with respect to a co-moving reference system.

We now consider a pressure force acting on a surface element s' that is at rest relative to S':

$$\begin{aligned} K'_x &= p's'\cdot\cos \ell' = p'\cdot s'_x \\ K'_y &= p'\cdot s'\cdot\cos m' = p'\cdot s'_y \\ K'_z &= p'\cdot s'\cdot\cos n' = p'\cdot s'_z \ , \end{aligned}$$

---

[1] This circumstance also justifies the procedure used in the preceding investigations, in which we introduced only interactions of a *purely electromagnetic nature* between the system considered and its surroundings. The results are valid in general.

where ℓ', m', n' denote the direction cosines of the normal (directed toward the interior of the body), and s'_x, s'_y, and s'_z the projections of s'. From equations (2) it follows that

$$\begin{aligned} s'_x &= s_x \\ s'_y &= \beta\cdot s_y \\ s'_z &= \beta\cdot s_z, \end{aligned}$$

where s_x, s_y, s_z are projections of the surface element with respect to S. For the components K_x, K_y, K_z of the pressure forces with respect to S, we therefore obtain from the last three systems of equations

$$K_x = K'_x = p's'_x = p'\cdot s_x = p'\cdot s \cos \ell$$

$$K_y = \frac{1}{\beta} K'_y = \frac{1}{\beta} p's'_y = p'\cdot s_y = p'\cdot s\cdot\cos m$$

$$K_z = \frac{1}{\beta} K'_z = \frac{1}{\beta} p's'_z = p'\cdot s_z = p'\cdot s\cdot\cos n$$

where s is the magnitude of the surface element, and ℓ, m, n denote the direction cosines of its normal with respect to S. We thus obtain the result that the pressure p' with respect to the co-moving system can be replaced with respect to another reference system by a pressure that has the same magnitude and is also perpendicular to the surface element. In our notation we thus have

$$p = p_0 \,. \tag{22}$$

Equations (16c), (29), and (22) enable us to determine the state of a physical system using the quantities E, V, p, which are defined with respect to the same system as the system's momentum G and velocity q, instead of using quantities E_0, V_0, p_0 referred to the co-moving reference system. E.g., if to a co-moving observer the state of the system under consideration is completely determined by two variables (V_0 and E_0), i.e., if the system's equation of state can be considered as a relation between p_0, V_0 and E_0, then, with the help of the above-mentioned equations, the equation of state can be brought to the form

$$\varphi(q,p,V,E) = 0 .$$

Similarly, equation (18c) can be brought to the form

[76] $$G = q\left\{\mu + \frac{E + pV}{c^2}\right\}, \tag{18d}$$

which, in combination with the equations expressing the principle of conservation of momentum

$$\frac{dG_x}{dt} = \Sigma\, K_x, \text{ etc.},$$

completely determines the translational motion of the system as a whole, if in addition to the quantities $\sum K_x$, etc. one also knows E, p, and V as functions of time, or if instead of the last three functions one knows three equivalent data regarding the conditions under which the motion of the system is taking place.

[77]

§14. *Examples*

Let the system under consideration consist of electromagnetic radiation, which is enclosed in a massless hollow body whose walls balance the radiation pressure. If no external forces act upon the hollow body, we may apply equations (16a) and (18a) to the entire system (including the hollow body). We will then have

$$E = \frac{E_0}{\sqrt{1 - \frac{q^2}{c^2}}}$$

$$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} E_0 = q \frac{E}{c^2} ,$$ [78]

where E_0 denotes the energy of the radiation with respect to a co-moving reference system.

However, if the walls of the hollow body are completely flexible and extensible, so that the radiation pressure exerted from within must be balanced by external forces exerted by bodies not belonging to the system considered, one has to apply equations (16c) and (18c) and insert into them the well-known value of the radiation pressure

$$p_0 = \frac{1}{3} \frac{E_0}{c^2} ,$$ [79]

so that one obtains

$$E = \frac{E_0\left(1 + \frac{1}{3} \frac{q^2}{c^2}\right)}{\sqrt{1 - \frac{q^2}{c^2}}}$$

$$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} \frac{\frac{4}{3} E_0}{c^2} .$$

Next we consider the case of an electrically charged massless body. If external forces do not act upon the body, we can once again apply formulas (16a) and (18a). Denoting the electrical energy relative to a co-moving reference system by E_0, we get

$$E = \frac{E_0}{\sqrt{1 - \frac{q^2}{c^2}}}$$

[80]

$$G = \frac{q}{\sqrt{1 - \frac{q^2}{c^2}}} \frac{\frac{4}{3} E_0}{c^2} .$$

A part of these values is to be allotted to the electromagnetic field, and the rest to the massless body that is subjected to forces due to its charge.[1]

§15. *The entropy and temperature of moving systems*

Of the variables that determine the state of a system, we have thus far only used pressure, volume, energy, velocity, and momentum, but have not yet dealt with thermal quantities. The reason for this was that for a system's *motion* it is irrelevant what kind of energy is supplied, so that we had no reason to distinguish between heat and mechanical work. However, we now want to introduce thermal quantities as well.

Let the state of a moving system be completely determined by the quantities q, V, and E. Obviously, in the case of such a system we have to consider as the heat supplied dQ the total energy increase minus the work produced by the pressure and that spent on increasing the momentum, so that we have

[82]

$$dQ = dE + pdV - qdG . \tag{23}$$

After having so defined the heat supplied for a moving system, we can introduce the absolute temperature T and the entropy η of the moving system by considering reversible cyclic processes in the same way it is done in textbooks of thermodynamics. For reversible processes the equation

---

[81] [1] Cf. A. Einstein, *Ann. d. Phys.* 23 (1907): 373-379.

$$dQ = Td\eta \tag{24}$$

is valid here too.

We now have to derive the equations relating the quantities dQ, η, T and the corresponding quantities dQ_0, η_0, T_0 which refer to a co-moving reference system. As far as entropy is concerned, I am repeating here the reasoning of Mr. Planck,[1] noting that the "primed" and "unprimed" reference systems should be understood as the reference systems S' and S.

"Let us imagine that the body is brought by some reversible, adiabatic process from a state in which it is at rest with respect to the unprimed system into a second state, in which it is at rest with respect to the primed reference system. If the body's entropy for the unprimed system in the initial state is denoted by η_1 and in the final state by η_2, then, because of the reversibility and adiabatic nature of the process, $\eta_1 = \eta_2$. But the process is reversible and adiabatic for the primed reference system as well, hence we will also have $\eta_2' = \eta_2'$."

"Now, if η_1' were not equal to η_1 but, say, $\eta_1' > \eta_1$, this would mean the following: The entropy of a body is larger for the reference system for which it is in motion than for the reference system for which it is at rest. [84] But this proposition would also require that $\eta_2' > \eta_2$, because in the latter state the body is at rest for the primed reference system while in motion for the unprimed one. [85] However, these two inequalities conflict with the two equalities established. Similarly, one cannot have $\eta_1' > \eta_1$; consequently $\eta_1' = \eta_1$ and, in general, $\eta' = \eta$, i.e., the entropy of the body does not depend on the choice of the reference system." [86]

Using our notation, we must therefore put

$$\eta = \eta_0. \tag{25}$$

If we now introduce the quantities E_0, p_0, and V_0 on the right-hand side of equation (23) by means of equations (16c), (18c), (20), and (22), we obtain

---

[1] M. Planck, "Zur Dynamik bewegter Systeme" [On the dynamics of moving systems]. *Sitzungsber. d. kgl. Preuss. Akad. d. Wissensch.* (1907). [83]

$$dQ = \sqrt{1 - \frac{q^2}{c^2}}\,(dE_0 + p_0 dV_0)$$

or

$$dQ = dQ_0 \cdot \sqrt{1 - \frac{q^2}{c^2}}\,. \qquad (26)$$

Further, since according to (24) the two equations

[87]
$$\begin{aligned} dQ &= Td\eta \\ dQ_0 &= T_0 d\eta_0 \end{aligned}$$

hold, one finally obtains, taking into account (25) and (26),

[88]
$$\frac{T}{T_0} = \sqrt{1 - \frac{q^2}{c^2}}\,. \qquad (27)$$

Thus, the temperature of a moving system is always lower with respect to a reference system that is in motion relative to it than with respect to a reference system that is at rest relative to it.

§16. *The dynamics of systems and the principle of least action*

In his treatise "On the dynamics of moving systems," Mr. Planck starts out from the principle of least action (and from the transformation equations for the pressure and temperature of black-body radiation)[1] and arrives at results that are identical with those established here. The question arises, therefore, as to how the foundations of his study and the present one are related.

We started from the energy principle and the principle of conservation of momentum. If the components of the resultant of the forces acting upon the system are called F_x, F_y, F_z, we can formulate in the following way the principles we used for reversible processes and a system whose state is defined by the variables q, V, T :

---

[89] [1] M. Planck, "Zur Dynamik bewegter Systeme" [On the dynamics of moving systems]. *Sitzungsber. d. kgl. Preuss. Akad. d. Wissensch.* (1907).

$$dE = F_x dx + F_y dy + F_z dz - pdV + TdS \tag{28}$$ [90]

$$F_x = \frac{dG_x}{dt}, \text{ etc.} \tag{29}$$

Keeping in mind that

$$F_x dx = F_x \dot{x}\, dt = \dot{x} dG = d(\dot{x}G_x) - G_x d\dot{x}\ , \text{ etc.}$$ [91]

and

$$Td\eta = d(T\eta) - \eta dT\ ,$$

one obtains from the above equations the relation

$$d(-E + T\eta + qG) = G_x d\dot{x} + G_y d\dot{y} + G_z d\dot{z} + pdV + \eta dT\ .$$

Since the right-hand side of this equation must also be a total differential, and taking into account (29), it follows that

$$\frac{d}{dt}\left[\frac{\partial H}{\partial \dot{x}}\right] = F_x \qquad \frac{d}{dt}\left[\frac{\partial H}{\partial \dot{y}}\right] = F_y \qquad \frac{d}{dt}\left[\frac{\partial H}{\partial \dot{z}}\right] = F_z$$

$$\frac{\partial H}{\partial V} = p \qquad \frac{\partial H}{\partial T} = \eta\ .$$

But these are the equations derivable by means of the principle of least action which Mr. Planck had used as his starting point. [92]

V. PRINCIPLE OF RELATIVITY AND GRAVITATION

§17. *Accelerated reference system and gravitational field*

So far we have applied the principle of relativity, i.e., the assumption that the physical laws are independent of the state of motion of the reference system, only to *nonaccelerated* reference systems. Is it conceivable that the principle of relativity also applies to systems that are accelerated relative to each other?

While this is not the place for a detailed discussion of this question, it will occur to anybody who has been following the applications of the principle of relativity. Therefore I will not refrain from taking a stand on this question here.

We consider two systems Σ_1 and Σ_2 in motion. Let Σ_1 be accelerated
[93] in the direction of its X-axis, and let γ be the (temporally constant) magnitude of that acceleration. Σ_2 shall be at rest, but it shall be located in a homogeneous gravitational field that imparts to all objects an acceleration $-\gamma$ in the direction of the X-axis.

[94] As far as we know, the physical laws with respect to Σ_1 do not differ from those with respect to Σ_2; this is based on the fact that all bodies are equally accelerated in the gravitational field. At our present state of experience we have thus no reason to assume that the systems Σ_1 and Σ_2 differ from each other in any respect, and in the discussion that follows, we shall therefore assume the complete physical equivalence of a gravitational field and a corresponding acceleration of the reference system.

This assumption extends the principle of relativity to the uniformly accelerated translational motion of the reference system. The heuristic value of this assumption rests on the fact that it permits the replacement of a homogeneous gravitational field by a uniformly accelerated reference system, the latter case being to some extent accessible to theoretical treatment.

§18. *Space and time in a uniformly accelerated reference system*

We first consider a body whose individual material points, at a given time t of the nonaccelerated reference system S, possess no velocity relative to S, but a certain acceleration. What is the influence of this acceleration γ on the shape of the body with respect to S?

If such an influence is present, it will consist of a constant-ratio dilatation in the direction of acceleration and possibly in the two directions perpendicular to it, since an effect of another kind is impossible for reasons of symmetry. The acceleration-caused dilatations (if such exist at all) must be even functions of γ; hence they can be neglected if one restricts oneself to the case in which γ is so small that terms of the second or higher power in γ may be neglected. Since we are going to restrict ourselves to that case, we do not have to assume that the acceleration has any influence on the shape of the body.

We now consider a reference system Σ that is uniformly accelerated relative to the nonaccelerated system S in the direction of the latter's X-axis. The clocks and measuring rods of Σ, examined at rest, shall be identical with the clocks and measuring rods of S. The coordinate origin of Σ shall move along the X-axis of S, and the axes of Σ shall be perpetually parallel to those of S. At any moment there exists a nonaccelerated reference system S' whose coordinate axes coincide with the coordinate axes of Σ at the moment in question (at a given time t' of S'). If the coordinates of a point event occurring at this time t' are ξ, η, ζ with respect to Σ, we will have

$$\left.\begin{aligned} x' &= \xi \\ y' &= \eta \\ z' &= \zeta \end{aligned}\right\},$$

because in accordance with what we said above, we are not to assume that acceleration affects the shape of the measuring instruments used for measuring ξ, η, ζ. We shall also imagine that the clocks of Σ are set at time t' of S' such that their readings at that moment equal t'. What about the rate of the clocks in the next time element τ?

First of all, we have to bear in mind that a specific effect of *acceleration* on the rate of the clocks of Σ need not be taken into account, since it would have to be of the order γ^2. Furthermore, since the effect of the velocity attained during τ on the rate of the clocks is negligible, and the distances traveled by the clocks during the time τ relative to those traveled by S' are also of the order τ^2, i.e., negligible, the readings of the clocks of Σ may be fully replaced by readings of the clocks of S' for the time element τ. [95]

From the foregoing it follows that, relative to Σ, light in vacuum is propagated during the time element τ with the universal velocity c if we define simultaneity in the system S' which is momentarily at rest relative to Σ, and if the clocks and measuring rods we use for measuring the time and length are identical with those used for the measurement of time and space in nonaccelerated systems. Thus the principle of constancy of the velocity of light can be used here too to define simultaneity if one restricts oneself to very short light paths.

We now imagine that the clocks of Σ are adjusted, in the way described, at that time $t = 0$ of S at which Σ is instantaneously at rest relative to S. The totality of readings of the clocks of Σ adjusted in [96] this way is called the "local time" σ of the system Σ. It is immediately evident that the physical meaning of the local time σ is as follows. If one uses the local time σ for the temporal evaluation of processes occurring in the individual space elements of Σ, then the laws obeyed by these processes cannot depend on the position of these space elements, i.e., on their coordinates, if not only the clocks, but also the other measuring tools used in the [97] various space elements are identical.

However, we must not simply refer to the local time σ as the "time" of Σ, because according to the definition given above, two point events occurring at different points of Σ are not simultaneous when their local times σ are equal. For if at time $t = 0$ two clocks of Σ are synchronous with respect to S and are subjected to the same motions, then they remain forever synchronous with respect to S. However, for this reason, in accordance with §4, they do not run synchronously with respect to a reference system S' instantaneously at rest relative to Σ but in motion relative to S, and hence according to our definition they do not run synchronously with respect to Σ either.

We now define the "time" τ of the system Σ as the totality of those readings of the clock situated at the coordinate origin of Σ which are, according to the above definition, simultaneous with the events which are to be temporally evaluated.[1]

We shall now determines the relation between the time τ and the local time σ of a point event. It follows from the first of equations (1) that

---

[1] Thus the symbol "τ" is used here in a different sense than above.

two events are simultaneous with respect to S', and thus also with respect to Σ, if

$$t_1 - \frac{v}{c^2} x_1 = t_2 - \frac{v}{c^2} x_2 ,$$

where the subscripts refer to the one or to the other point event, respectively. We shall first confine ourselves to the consideration of times that are so short[1] that all terms containing the second or higher power of τ or v can be omitted; taking (1) and (29) into account, we then have to put [98]

$$x_2 - x_1 = x_2' - x_2' = \xi_2 - \xi_1$$
$$t_1 = \sigma_1 \qquad t_2 = \sigma_2$$
$$v = \gamma t = \gamma\tau ,$$ [99]

so that we obtain from the above equation

$$\sigma_2 - \sigma_1 = \frac{\gamma\tau}{c^2} (\xi_2 - \xi_1) .$$

If we move the first point event to the coordinate origin, so that $\sigma_1 = \tau$ and $\xi_1 = 0$, we obtain, omitting the subscript for the second point event,

$$\sigma = \tau\left[1 + \frac{\gamma\xi}{c^2}\right] . \qquad (30)$$

This equation holds first of all if τ and ξ lie below certain limits. It is obvious that it holds for arbitrarily large τ if the acceleration γ is constant with respect to Σ, because the relation between σ and τ must then be linear. Equation (30) does not hold for arbitrarily large ξ. From the fact that the choice of the coordinate origin must not affect the relation, one must conclude that, strictly speaking, equation (30) should be replaced by the equation

$$\sigma = \tau e^{\frac{\gamma\xi}{c^2}} .$$

Nevertheless, we shall maintain formula (30).

---

[1] In accordance with (1), we thereby also assume a certain restriction with respect to the values of $\xi = x'$.

According to §17, equation (30) is also applicable to a coordinate system in which a homogeneous gravitational field is acting. In that case we have to put $\Phi = \gamma\xi$, where Φ is the gravitational potential, so that we obtain

$$\sigma = \tau\left[1 + \frac{\Phi}{c^2}\right] . \tag{30a}$$

We have defined two kinds of times for Σ. Which of the two definitions do we have to use in the various cases? Let us assume that at two locations of different gravitational potentials $(\gamma\xi)$ there exists one physical system each, and we want to compare their physical quantities. To do this, the most natural procedure might be as follows: First we take our measuring tools to the first physical system and carry out our measurements there; then we take our measuring tools to the second system to carry out the same measurement here. If the two sets of measurements give the same results, we shall denote the two physical systems as "equal." The measuring tools include a clock with which we measure local times σ. From this it follows that to define the physical quantities at some position of the gravitational field, it is natural to use the time σ.

However, if we deal with a phenomenon in which objects situated at positions with different gravitational potentials must be considered simultaneously, we have to use the time τ in those terms in which time occurs explicitly (i.e., not only in the definition of physical quantities), because otherwise the simultaneity of the events would not be expressed by the equality of the time values of the two events. Since in the definition of the time τ a clock situated in an arbitrarily chosen position is used, but not an arbitrarily chosen instant, when using time τ the laws of nature can vary with position but not with time.

§19. *The effect of the gravitational field on clocks*

If a clock showing local time is located in a point P of gravitational potential Φ, then, according to (30a), its reading will be $(1 + \frac{\Phi}{c^2})$ times greater than the time τ, i.e., it runs $(1 + \frac{\Phi}{c^2})$ times faster than an identical clock located at the coordinate origin. Suppose an observer located somewhere in space perceives the indications of the two clocks in a certain way, e.g., optically. As the time $\Delta\tau$ that elapses between the instants at which a clock indication occurs and at which this indication is perceived by the observer is independent of τ, for an observer situated somewhere in space the clock in point P runs $(1 + \frac{\Phi}{c^2})$ times faster than the clock at the coordinate origin. In this sense we may say that the process occurring in the clock, and, more generally, any physical process, proceeds faster the greater the gravitational potential at the position of the process taking place.

There exist "clocks" that are present at locations of different gravitational potentials and whose rates can be controlled with great precision; these are the producers of spectral lines. It can be concluded from the aforesaid[1] that the wave length of light coming from the sun's surface, which originates from such a producer, is larger by about one part in two millionth than that of light produced by the same substance on earth. [100]

§20. *The effect of gravitation on electromagnetic phenomena*

If we refer an electromagnetic process at some point of time to a non-accelerated reference system S' that is instantaneously at rest relative to the reference system Σ accelerated as above, then the following equations will hold according to (5) and (6):

$$\frac{1}{c}\left[\rho' u'_x + \frac{\partial X'}{\partial t'}\right] = \frac{\partial N'}{\partial y'} - \frac{\partial M'}{\partial z'}\text{, etc.}$$

and

$$\frac{1}{c}\frac{\partial L'}{\partial t'} = \frac{\partial Y'}{\partial z'} - \frac{\partial Z'}{\partial y'}\text{, etc.}$$

In accordance with the above, we may readily equate the S'-referred quantities ρ', u', X', L', x', etc., with the corresponding Σ-referred

---

[1] While assuming that equation (30a) holds for an inhomogeneous gravitational field as well.

quantities ρ, u, X, L, ξ, etc., if we limit ourselves to an infinitesimally short period[1] that is infinitesimally close to the time of relative rest of S' and Σ. Further, we have to replace t' by the local time σ. However, we must not simply put

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial \sigma} ,$$

because a point which is at rest relative to Σ, and to which equations transformed to Σ should refer, changes its velocity relative to S' during the time element $dt' = d\sigma$, to which change, according to equations (7a) and (7b), there corresponds a temporal change of the Σ-related field component. Hence we have to put

$$\frac{\partial X'}{\partial t'} = \frac{\partial X}{\partial \sigma} \qquad \frac{\partial L'}{\partial t'} = \frac{\partial L}{\partial \sigma}$$

$$\frac{\partial Y'}{\partial t'} = \frac{\partial Y}{\partial \sigma} + \frac{\gamma}{c} N \qquad \frac{\partial M'}{\partial t'} = \frac{\partial M}{\partial \sigma} - \frac{\gamma}{c} Z$$

$$\frac{\partial Z'}{\partial t'} = \frac{\partial Z}{\partial \sigma} - \frac{\gamma}{c} M \qquad \frac{\partial N'}{\partial t'} = \frac{\partial N}{\partial \sigma} + \frac{\gamma}{c} Y .$$

Hence the Σ-referred electromagnetic equations are

$$\frac{1}{c}\left[\rho \mathrm{u}_\xi + \frac{\partial X}{\partial \sigma}\right] = \frac{\partial N}{\partial \eta} - \frac{\partial M}{\partial \zeta}$$

$$\frac{1}{c}\left[\rho \mathrm{u}_\eta + \frac{\partial Y}{\partial \sigma} + \frac{\gamma}{c} N\right] = \frac{\partial L}{\partial \zeta} - \frac{\partial N}{\partial \xi}$$

$$\frac{1}{c}\left[\rho \mathrm{u}_\xi + \frac{\partial Z}{\partial \sigma} - \frac{\gamma}{c} M\right] = \frac{\partial M}{\partial \xi} - \frac{\partial L}{\partial \eta}$$

$$\frac{1}{c}\frac{\partial L}{\partial \sigma} = \frac{\partial Y}{\partial \zeta} - \frac{\partial Z}{\partial \eta}$$

$$\frac{1}{c}\left[\frac{\partial M}{\partial \sigma} - \frac{\gamma}{c} Z\right] = \frac{\partial Z}{\partial \xi} - \frac{\partial X}{\partial \zeta}$$

$$\frac{1}{c}\left[\frac{\partial N}{\partial \sigma} + \frac{\gamma}{c} Y\right] = \frac{\partial X}{\partial \eta} - \frac{\partial Y}{\partial \xi} .$$

---

[1] This restriction does not affect the range of validity of our results because inherently the laws to be derived cannot depend on the time.

We multiply these equations by $\left[1 + \frac{\gamma\xi}{c^2}\right]$ and put for the sake of brevity

$$X^* = X\left[1 + \frac{\gamma\xi}{c^2}\right], \qquad Y^* = Y\left[1 + \frac{\gamma\xi}{c^2}\right] \text{ , etc.}$$

$$\rho^* = \rho\left[1 + \frac{\gamma\xi}{c^2}\right] .$$

Neglecting terms of the second power in γ, we obtain the equations

$$\left.\begin{aligned} \frac{1}{c}\left[\rho^* u_\xi + \frac{\partial X^*}{\partial\sigma}\right] &= \frac{\partial N^*}{\partial\eta} - \frac{\partial M^*}{\partial\zeta} \\ \frac{1}{c}\left[\rho^* u_\eta + \frac{\partial Y^*}{\partial\sigma}\right] &= \frac{\partial L^*}{\partial\zeta} - \frac{\partial N^*}{\partial\xi} \\ \frac{1}{c}\left[\rho^* u_\zeta + \frac{\partial Z^*}{\partial\sigma}\right] &= \frac{\partial M^*}{\partial\xi} - \frac{\partial L^*}{\partial\eta} \end{aligned}\right\} \tag{31a}$$

$$\left.\begin{aligned} \frac{1}{c}\frac{\partial L^*}{\partial\sigma} &= \frac{\partial Y^*}{\partial\zeta} - \frac{\partial Z^*}{\partial\eta} \\ \frac{1}{c}\frac{\partial M^*}{\partial\sigma} &= \frac{\partial Z^*}{\partial\xi} - \frac{\partial X^*}{\partial\zeta} \\ \frac{1}{c}\frac{\partial N^*}{\partial\sigma} &= \frac{\partial X^*}{\partial\eta} - \frac{\partial Y^*}{\partial\xi} . \end{aligned}\right\} \tag{32a}$$

These equations show first of all how the gravitational field affects the static and stationary phenomena. The same laws hold as in the gravitation-free field, except that the field components X, etc. are replaced by $X\left[1 + \frac{\gamma\xi}{c^2}\right]$, etc., and ρ is replaced by $\rho\left[1 + \frac{\gamma\xi}{c^2}\right]$.

Furthermore, to follow the development of nonstationary states, we make use of the time τ in the terms differentiated with respect to time as well as in the definition of the velocity of electricity, i.e., we put according to (30)

$$\frac{\partial}{\partial\tau} = \left[1 + \frac{\gamma\xi}{c^2}\right]\frac{\partial}{\partial\tau} \qquad [101]$$

and

$$w_\xi = \left[1 + \frac{\gamma\xi}{c^2}\right] . \qquad [102]$$

We thus obtain

$$\frac{1}{c\left[1+\frac{\gamma\xi}{c^2}\right]}\left[\rho^* w_\xi + \frac{\partial X^*}{\partial \tau}\right] = \frac{\partial N^*}{\partial \eta} - \frac{\partial M^*}{\partial \zeta} \quad \text{etc.} \tag{31b}$$

and

[103]

$$\frac{1}{c\left[1+\frac{\gamma\xi}{c^2}\right]}\frac{\partial L^*}{\partial \tau} = \frac{\partial Y^*}{\partial \zeta} = \frac{\partial Z^*}{\partial \eta} \quad \text{etc.} \tag{32b}$$

These equations too have the same form as the corresponding equations of the nonaccelerated or gravitation-free space; however, c is here replaced by the value

$$c\left[1+\frac{\gamma\xi}{c^2}\right] = c\left[1+\frac{\Phi}{c^2}\right] .$$

From this it follows that those light rays that do not propagate along the ξ-axis are bent by the gravitational field; it can easily be seen that the change of direction amounts to $\frac{\gamma}{c^2}\sin\varphi$ per cm light path, where φ
[104] denotes the angle between the direction of gravity and that of the light ray.

With the help of these equations and the equations relating the field strength and the electric current of one point, which are known from the optics of bodies at rest, we can calculate the effect of the gravitational field on optical phenomena in bodies at rest. One has to bear in mind, however, that the above-mentioned equations from the optics of bodies at rest hold for the local time σ. Unfortunately, the effect of the terrestrial gravitational field is so small according to our theory (because of the smallness of $\frac{\gamma x}{c^2}$) that there is no prospect of a comparison of the results of
[105] the theory with experience.

If we successively multiply equations (31a) and (32a) by $\frac{X^*}{4\pi} \cdots\cdots \frac{N^*}{4\pi}$ and integrate over infinite space, we obtain, using our earlier notation,

[106]

$$\int\left[1+\frac{\gamma\xi}{c^2}\right]^2\frac{\rho}{4\pi}(uX + u_\eta Y + uZ)\,d\omega + \int\left[1+\frac{\gamma\xi}{c^2}\right]^2\cdot\frac{1}{8\pi}\frac{\partial}{\partial\sigma}(X^2 + Y^2 + \cdots + N^2)\,d\omega = 0 .$$

[107] $\frac{\rho}{4\pi}(uX + u_\eta Y + u_\xi Z)$ is the energy η_σ supplied to the matter per unit volume and unit local time σ if this energy is measured by measuring tools situated at the corresponding location. Hence, according to (30),

$\eta_\tau = \eta^\sigma\left[1 + \frac{\gamma\xi}{c^2}\right]$ is the (similarly measured) energy supplied to the matter per [108] unit volume and unit local time τ; $\frac{1}{8\pi}(X^2 + Y^2 \cdots + N^2)$ is the electromagnetic energy ϵ per unit volume, measured the same way. If we take into account that according to (30) we have to set $\frac{\partial}{\partial\sigma} = \left[1 - \frac{\gamma\xi}{c^2}\right]\frac{\partial}{\partial\tau}$, we obtain

$$\int \left[1 + \frac{\gamma\xi}{c^2}\right]\eta_\tau d\omega + \frac{d}{d\tau}\left\{\int \left[1 + \frac{\gamma\xi}{c^2}\right]\epsilon \, d\omega\right\} = 0 \ .$$

This equation expresses the principle of conservation of energy and contains a very remarkable result. An energy, or energy input, that, measured locally, has the value $E = \epsilon d\omega$ or $E = \eta \, d\omega d\tau$, respectively, contributes to the energy integral, in addition to the value E that corresponds to its magnitude, also a value $\frac{E}{c^2}\gamma\xi = \frac{E}{c^2}\Phi$ that corresponds to its *position*. Thus, to each energy E in the gravitational field there corresponds an energy of position that equals the potential energy of a "ponderable" mass of magnitude $\frac{E}{c^2}$.

Thus the proposition derived in §11, that to an amount of energy E there corresponds a mass of magnitude $\frac{E}{c^2}$, holds not only for the *inertial* but also for the *gravitational* mass, if the assumption introduced in §17 is correct.

(Received on 4 December 1907)

Doc. 48

A NEW ELECTROSTATIC METHOD FOR THE MEASUREMENT OF SMALL QUANTITIES OF ELECTRICITY

by A. Einstein

[*Physikalische Zeitschrift* 9 (1908): 216-217]

As we know, electric potentials as low as about 10^{-6} volts can be measured by sensitive electrostatic quadrant electrometers if a sufficiently [1] great auxiliary potential is supplied to the needle. However, an increase in this auxiliary potential results in a decrease, rather than increase, in the sensitivity of the instrument if one measures *quantities of electricity*. The higher the potential of the needle, the smaller the deflection caused by a given quantity of electricity. If the absolute value of the potential of the needle is large compared with the potential difference between the quadrants, then the deflection depends only on the product of the potential and the quantity of electricity supplied, and thus on the electric energy supplied, [2] and the energy required for the deflection must be taken from the energy of the system being measured. This circumstance determines the limit actually achievable of the sensitivity of the quadrant electrometer and of analogous instruments when measuring quantities of electricity or energy, respectively.

However, it is possible to construct measuring instruments in which the energy required for deflection is not taken from the system being measured, but from an auxiliary source, so that it will be possible to exceed the actual sensitivity limit mentioned above. In the following, I describe the scheme of [3] an induction machine with which, in my opinion, this goal can be achieved.

Let A_1 and A_1' (Fig. 1) be two fixed conductors, along which shall slide two rigidly connected metal leaflets B, fastened, for example, to a small wheel. The leaflets have two contact pins b, which are within the range of fixed contact springs K_1 and K_1', respectively. Let K_1 be grounded and K_1' conductively connected with A_1'.

A_1 shall be kept permanently at a positive potential P_1. When the leaflet that is just passing touches K_1, the electric charge on A_1 induces an opposite electric charge $-e$ on b. Once this leaflet is opposite A_1', so that it touches K_1', it delivers negative electricity to A_1'. In such a way

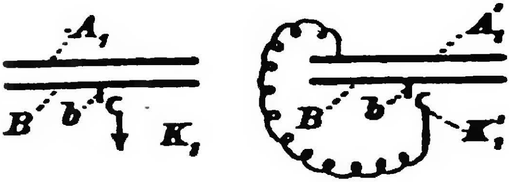


Fig. 1

each passing leaflet will change the quantity of electricity on $A_1^{\prime}$ until a stationary state has been reached. If $P_1^{\prime}$ denotes the absolute value of the negative potential of $A_1^{\prime}$ in the stationary state, then we must have

$$\frac{P_1^{\prime}}{P_1} = a_1 ,$$

where a_1 is the transformation ratio, a constant independent of P_1. If A_1 and $A_1^{\prime}$ are leaflet-shaped, a_1 will be a proper fraction. But if A_1 and $A_1^{\prime}$ are shaped like bows, which at the instants of contact formation surround the leaflets B on both sides, we can easily arrange that $a_1 > 1$, e.g., $a_1 = 10$. This is what we will assume in the following.

Let us envisage several such elements connected in series according to the accompanying scheme (Fig. 2). The secondary conductor $A_1^{\prime}$ of the first element shall be connected with the primary one A_2 of the second element, the secondary conductor $A_2^{\prime}$ of the second element with the primary one A_3 of the third element, etc. The secondary conductor of the last element shall be connected to an electrometer V.

If one imagines a certain potential P_1 supplied to the primary conductor A_1, then a stationary state of the whole arrangement will be established after a certain time. The following equation will then hold for it:

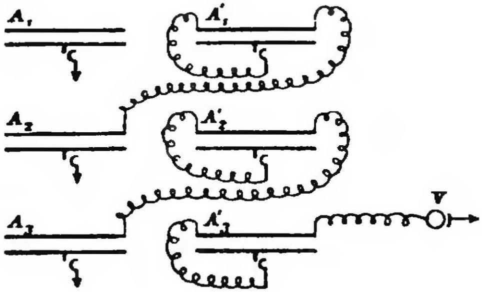


Fig. 2

$$P_2 = P_1' = P_1 \cdot a_1$$
$$P_2 = P_2' = P_2 \cdot a_2 = P_1 \cdot a_1 \cdot a_2$$
$$P_3' = P_3 \cdot a_3 = P_1 \cdot a_1 \cdot a_2 \cdot a_3 .$$

If there are n elements, all of them with the same transmission ratio a, then we will have

$$P_n' = P_1 \cdot a^n \ .$$

One can see that the energy that has to be supplied to the mobile system of the electrometer is taken from the mechanical energy imparted to the leaflets B, and not from the system to be measured, which is connected to the leaflet A. The sensitivity of the procedure is limited only by the external [4] sources of error, since a^n can be made arbitrarily large by increasing n.

Of course, the whole arrangement can be made to be bipolar. In that case each element gets two primary and two secondary conductors.

Since an increase in the sensitivity of electrostatic methods of measurement is of importance for the study of radioactivity, I hope that some [5] physicist will become interested in this matter. I would gladly inform him [6] about my further considerations on this subject. I was led to the plan presented herein by thinking about how the spontaneous charging of conductors[1] required by the molecular theory of heat, which is analogous to Brownian motion, can be detected and measured. I hope that, with the plan described, I have brought this problem too one step closer to its solution.

Bern, 13 February 1908. (Received on 15 February 1908)

---

[1] A. Einstein, *Ann. d. Phys.* 22 (1907): 569.

Doc. 49

CORRECTIONS TO THE PAPER: "ON THE RELATIVITY PRINCIPLE AND THE CONCLUSIONS DRAWN FROM IT[1]"

by A. Einstein

[*Jahrbuch der Radioaktivität und Elektronik* 5 (1908): 98-99]

During the proofreading of the article cited I missed unfortunately several errors that have to be corrected because they impede the reading of [2] the article.

Formula 15b (p. 435) should read

$$\frac{d}{dt}\left[\int\frac{1}{4\pi c}(YN - ZM)\,d\omega\right] + \Sigma\frac{\mu\dot{x}}{\sqrt{1-\frac{q^2}{c^2}}} = 0 .$$

The factor $\frac{4}{3}$ in the second formula on p. 451 is in error: the formula should read

$$G = \frac{q}{\sqrt{1-\frac{q^2}{c^2}}}\frac{E_0}{c^2} .$$

Formula 28 on p. 453 should read

$$dE = F_x dx + F_y dy + F_z dz - pdV + Td\eta .$$

A few lines further on, the subscript in G_x should be added. In the penultimate line on p. 455 it should read "replaceable" instead of "usable." [Translator's note: This correction does not apply to the translated version.]

On p. 451 it should read

---

[1] [1] This *Jahrbuch* 4 (1907): 411.

$$\frac{\partial}{\partial\tau} = \left[1 + \frac{\gamma\xi}{c^2}\right] \frac{\partial}{\partial\sigma}$$

and

$$w_\xi = \left[1 + \frac{\gamma\xi}{c^2}\right] u_\xi \ .$$

On p. 462 the subscripts in the quantities u_ξ and u_ζ have to be added. Also, in about the middle of this page a mistake in sign should be corrected: the equation should read

$$\eta_\sigma = \eta_\tau\left[1 - \frac{\gamma\xi}{c^2}\right] \ .$$

A letter by Mr. Planck induced me to add the following supplementary remark so as to prevent a misunderstanding that could arise easily:

In the section "Principle of relativity and gravitation", a reference system at rest situated in a temporally constant, homogeneous gravitational field is treated as physically equivalent to a uniformly accelerated, gravitation-free reference system. The concept "uniformly accelerated" needs further clarification.

If—as in our case—one considers a rectilinear motion (of the system Σ), the acceleration is given by the expression $\frac{dv}{dt}$, where v denotes the velocity. According to the kinematics in use up to now, $\frac{dv}{dt}$ is independent of the state of motion of the (nonaccelerated) reference system, so that one might speak directly of (instantaneous) acceleration when the motion in a certain time element is given. According to the kinematics used by us, $\frac{dv}{dt}$ does depend on the state of motion of the (nonaccelerated) reference system. But among all the values of acceleration that can be so obtained for a certain motion epoch, that one is distinguished which corresponds to a reference system with respect to which the body considered has the velocity $v = 0$. It is this value of acceleration which has to remain constant in our "uniformly [3] accelerated" system. The relation $v = \gamma t$ used on p. 457 thus holds only in first approximation; however, this is sufficient, because only terms linear in t and τ, respectively, have to be taken into account in these considerations.

(Received on 3 March 1908)

Doc. 50

ELEMENTARY THEORY OF BROWNIAN[1] MOTION

by A. Einstein

[*Zeitschrift für Elektrochemie und angewandte physikalishe Chemie* 14 (1908): 235-239]

[2] In a conversation, Professor R. Lorenz pointed out to me that many chemists would welcome an elementary theory of Brownian motion. Responding to his request, I present in the following a simple theory of this phenomenon. The train of thought to be conveyed, in brief, is as follows: First we investigate how the process of diffusion in an undissociated dilute solution depends on the distribution of the osmotic pressure in the solution and on the mobility of the dissolved matter relative to the solvent. For the case that a molecule of the dissolved matter is large compared with a molecule of the solvent we thus obtain an expression for the coefficient of diffusion in which no quantities appear which depend on the nature of the solvent other than the viscosity of the solvent and the diameter of the dissolved molecules.

Then we attribute the process of diffusion to the random motions of the dissolved molecules and find out how the mean magnitude of these random [3] motions of the dissolved molecules can be calculated from the coefficient of diffusion, i.e., according to the result mentioned above, from the viscosity of the solvent and the size of the dissolved molecules. The result thus obtained is then valid not only for true dissolved molecules but also for any small corpuscules suspended in the liquid.

§1. *Diffusion and osmotic pressure*

Let the cylindric vessel Z (Fig. 93) be filled with a dilute solution. Let, further, the interior of Z be divided in two parts A and B by the movable piston K, which constitutes a semipermeable wall. If the concentration

---

[1] By Brownian motion we understand the irregular motion performed by microscopically small particles suspended in a liquid. Cf., e.g., The Svedberg, *Zeitsch. f. Elektrochemie* 12 (1906): pp. 47 and 51. [1]

of the solution is higher in A than in B, then an external force, directed to the left, must be applied to the piston to maintain it in equilibrium, and this force is equal to the difference between the two osmotic pressures exerted by the dissolved substance on the piston from the left and from the right, respectively. If this external pressure is not applied to the piston, the latter will move to the right under the influence of the greater osmotic pressure exerted by the solution in A until the concentrations in A and B no longer differ. This consideration demonstrates that it is precisely the forces of osmotic pressure which cause the equalization of concentrations in diffusion; because we can prevent diffusion, i.e., the equalization of concentrations, by counterbalancing the osmotic differences, which correspond to differences in concentrations, with external forces acting on semipermeable walls. It has been known for a long time that osmotic pressure can be considered as the motive force in processes of diffusion. As we know, Nernst used this as the basis for his investigation on the connection between ionic mobility, the coefficient of diffusion, and the EMF in concentration cells. [4]

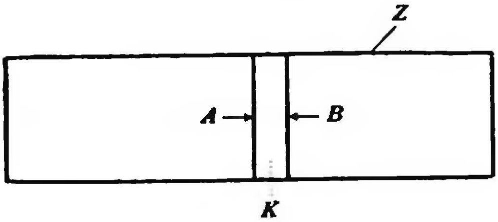


Fig. 93

Let diffusion take place along the cylinder's axis inside the cylinder Z (Fig. 94), whose cross section shall be = 1. Let us first examine the osmotic forces causing the diffusional motion of the dissolved substance contained between the infinitesimally close planes E and E'. From the left, the force of osmotic pressure p acts on the lamina's boundary surface E, and from the right, the pressure p' acts on the boundary surface E'; the resultant of the pressure forces is therefore

$$p - p'.$$

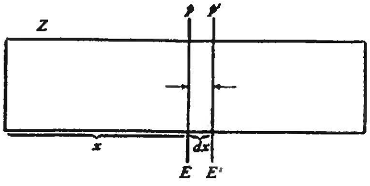


Fig. 94

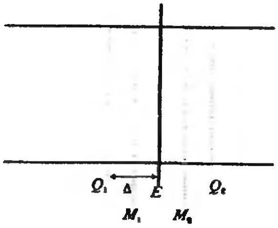


Fig. 95

We will denote the distance of the surface E from the left end of the vessel by x, and the distance of the surface E' from the same end of the vessel by $x + dx$; then dx also equals the volume of the liquid lamina considered. Since $p - p'$ is the osmotic pressure acting on the volume dx of the dissolved substance,

$$K = \frac{p - p'}{dx} = - \frac{p' - p}{dx} = - \frac{dp}{dx}$$

is the osmotic force acting on the dissolved substance contained in the unit
[5] volume. Since, further, the osmotic pressure is given by the equation

$$p = RT\nu,$$

where R denotes the constant of the gas equation ($8.31 \cdot 10^7$), T the absolute temperature, and ν the number of dissolved gram-molecules per unit volume, we get, finally, the following expression for the osmotic force K acting on the dissolved substance per unit volume

$$K = - RT \frac{d\nu}{dx} . \tag{1}$$

To be able to calculate the diffusional motions that these motive forces can produce, we must also know how great a resistance is offered by the solvent to the motion of the dissolved substance. If a motive force k acts on a molecule, it imparts to it a proportional velocity v according to the equation

$$v = \frac{k}{\mathfrak{R}}, \tag{2}$$

where $\mathfrak{R}$ is a constant which we will call the frictional resistance of the molecule. In general, this frictional resistance cannot be determined by theoretical methods. But if we are allowed to regard the molecule approximately as a sphere that is large compared with a molecule of the solvent, then we can determine the frictional resistance of the dissolved molecule by the methods of ordinary hydrodynamics, in which the molecular constitution of the liquid is not taken into account. Within the limits of validity of ordinary hydrodynamics, a sphere moving in a liquid obeys equation (2), where we put

$$\mathfrak{R} = 6\pi\eta\rho. \tag{3}$$

Here η denotes the coefficient of viscosity of the liquid, and ρ the radius of the sphere. If we can assume that the molecules of a dissolved substance are approximately spherical and large compared with the molecules of the solvent, then equation (3) may be applied to the individual dissolved molecules.

Now we can calculate the amount of dissolved substance diffusing through a cross section of the cylinder per unit time. The unit volume contains ν gram-molecules, which amounts to νN real molecules, where N denotes the number of real molecules in one gram-molecule. If a force K is distributed over these νN molecules contained in the unit volume, it will impart a velocity to them that is νN times smaller than the velocity it would be able to impart to a single molecule if it acted upon the latter alone. Taking into account equation (2), we get therefore for the velocity v that the force K can impart to νN molecules

$$v = \frac{1}{\nu N} \cdot \frac{K}{\mathfrak{R}}.$$

In the case considered, K is equal to the osmotic force exerted on the νN molecules contained in the unit volume, which we determined before, so that we get from this, using equation (1),

(4) $$v\nu = -\frac{RT}{N}\cdot\frac{1}{\mathfrak{R}}\cdot\frac{d\nu}{dx}\,.$$

The left-hand side contains the product of the concentration ν of the dissolved substance and the velocity with which the dissolved substance is moved along by the process of diffusion. This product represents therefore the quantity of the dissolved substance (in gram-molecules) transported per second through unit cross section by diffusion. The factor of $\frac{d\nu}{dx}$ on the right-hand side of this equation is therefore[1] nothing other than the coefficient of diffusion D of the solution considered. Hence, we have in general

(5) $$D = \frac{RT}{N}\cdot\frac{1}{\mathfrak{R}}\,,$$

and, in case the diffusing molecules can be considered as spherical and large compared with the molecules of the solvent, we have according to equation (3)

(5a) $$D = \frac{RT}{N}\cdot\frac{1}{6\pi\eta\rho}\,.$$

Thus, in the case just mentioned, the coefficient of diffusion does not depend on any constants characteristic for the substances in question other than the viscosity of the solvent η and the radius ρ of the molecule.[2]

---

[1] It should be noted that the numerical value of the coefficient of diffusion is independent of the choice of the unit for the concentration.

[2] This equation permits the approximate determination of the radius of (large) molecules from the coefficient of diffusion if the latter is know, since

$$\rho = \frac{RT}{6\pi N\eta}\cdot\frac{1}{D}\,,$$

where we have to put $R = 8.31\cdot 10^7$ and $N = 6\cdot 10^{23}$. To be sure, the value of
[6] N has a margin of uncertainty of about 50%. This relationship might be of significance for the determination of the approximate size of molecules in
[7] colloidal solutions.

§2. *Diffusion and random motion of molecules*

The molecular theory of heat affords yet another standpoint from which the process of diffusion can be viewed. The process of random motion, which is what the heat content of a substance must be considered to be, will cause the individual molecules of a liquid to change their position in the most random manner imaginable. This, as it were, haphazard meandering of the molecules of the dissolved substance in a solution will have as a consequence that the initial nonuniform distribution of concentration will gradually give way to a uniform one.

We will now consider this process in somewhat greater detail, limiting ourselves again to the case considered in §1, where only diffusion in one single direction, namely in the direction of the axis (x-axis) of the cylinder Z has to be taken into account. We imagine that we know the x-coordinates of all dissolved molecules at a certain time t, and also at time $t + \tau$, where τ denotes a time interval so short that the concentrations in our solution change very little during it. During this time τ, the x-coordinate of the first dissolved molecule will change by a certain quantity Δ_1 on account of the random thermal motion, that of the second molecule will change by Δ_2, etc. These displacements Δ_1, Δ_2, etc., will be in part negative (directed to the left) and in part positive (directed to the right). Furthermore, the magnitude of these displacements will vary from molecule to molecule. But since we assume, as before, that the solution is dilute, this displacement is determined only by the surrounding solvent, while the rest of the dissolved molecules has no appreciable effect; for that reason, these displacements Δ will *on the average* be of equal magnitude in parts of the solution having differing concentrations, and will be just as often positive as negative.

We now want to see how much of the substance diffuses through the unit cross section of our solution during time τ if we know the magnitude of the displacements Δ in the direction of the cylinder axis experienced on the average by the dissolved molecules. To simplify this consideration, we will assume that all molecules undergo an equally large displacement Δ, with half of the molecules undergoing the displacement $+\Delta$, (i.e., to the right), and the other half the displacement $-\Delta$ (i.e., to the left). We thus replace the individual displacements Δ_1, Δ_2, etc., by their mean value Δ. [8]

According to our simplifying assumption, the plane E of our cylinder (Fig. 95) [on p. 320] can be crossed during time τ from left to right only by those dissolved molecules which prior to interval τ were located left of E, at a distance from E smaller than Δ. These molecules are all located between the planes Q_1 and E (Fig. 95). But since only half of these molecules experience the displacement $+\Delta$, only half of them will cross the plane E. But one-half of the dissolved substance contained between Q_1 and E amounts, in gram-molecules, to

$$\frac{1}{2} \nu_1 \Delta,$$

where ν_1 denotes the mean concentration in the volume Q_1E, i.e., the concentration in the midplane M_1. Since the cross section equals 1, Δ represents the volume enclosed between Q_1 and E, which, when multiplied by the mean concentration, gives the dissolved substance contained in this volume in gram-molecules.

By an analogous consideration we find that the amount of dissolved substance crossing E from right to left during time τ equals

$$\frac{1}{2} \nu_2 \Delta,$$

where ν_2 denotes the concentration in the midplane M_2. The amount of substance diffusing through E from left to right during time τ is obviously equal to the difference of these two values, and hence equals

$$\frac{1}{2}\Delta(\nu_1 - \nu_2). \tag{6}$$

ν_1 and ν_2 are the concentrations in two cross sections separated by the very small distance Δ. If we again denote a cross section's distance from the left end of the cylinder by x, we will have according to the definition of the differential quotient

$$\frac{\nu_2 - \nu_1}{\Delta} = \frac{d\nu}{dx} ,$$

and from this

$$\nu_1 - \nu_2 = -\Delta \frac{d\nu}{dx} ,$$

so that the amount of substance diffusing during τ through E equals

(6a) $$-\frac{1}{2}\Delta^2 \frac{d\nu}{dx} .$$

The amount of substance, expressed in gram-molecules, diffusing through E in *unit time* hence equals

$$-\frac{1}{2} \frac{\Delta^2}{\tau} \frac{d\nu}{dx} .$$

With this we have obtained a second value for the coefficient of diffusion D. We have

(7) $$D = \frac{1}{2} \frac{\Delta^2}{\tau} ,$$

where Δ denotes the path travelled on the average[1] by a dissolved molecule during time τ in the direction of the x-axis.

Solving (7) for Δ, we obtain

(7a) $$\Delta = \sqrt{2D}\sqrt{\tau}.$$ [10]

§3. *Motion of individual molecules. Brownian motion*

If we equate the values for the diffusion coefficient in equations (5) and (7), we obtain by solving for Δ

---

[1] To be more precise, Δ equals the square root of the mean of the squares of the individual displacements Δ_1^2, Δ_2^2, etc. For greater accuracy, we should therefore write $\sqrt{\Delta^2}$ instead of Δ. [9]

(8)
$$\Delta = \sqrt{\frac{2RT}{N\mathfrak{R}}}\sqrt{\tau}\ .$$

We see from this formula that the path travelled on the average by a molecule is not proportional to the time[1], but to the square root of the time. This is due to the fact that the paths travelled in two consecutive time units are not always to be added, but just as frequently are to be subtracted. The displacement experienced on the average by a molecule on account of random molecular motion can be calculated according to equation (7a) from the coefficient of diffusion, or according to equation (8) from the force of resistance $\mathfrak{R}$ offered to a forced motion proceeding with velocity $v = 1$.

If the dissolved molecule is spherical and large compared with the molecule of the solvent, we can substitute for $\mathfrak{R}$ in equation (8) the value given in equation (3), so that we get

(8a)
$$\Delta = \sqrt{\frac{RT}{N}\cdot\frac{1}{3\pi\eta\rho}}\sqrt{\tau}.$$

This equation permits us to calculate the displacement average[2] Δ from the temperature T, the viscosity of the solvent η, and the molecular radius ρ.

But according to the molecular-kinetic concept, there exists no fundamental difference between a dissolved molecule and a suspended corpuscle. We must therefore consider equation (8a) to be valid for any kind of suspended spherical particles as well.

We now calculate the path Δ travelled on the average by a particle with a diameter of 1 micron in 1 second in a particular direction in water at room temperature. We put

$R = 8.31\cdot10^7$, $\eta = 0.0135$,
[12] $T = 290$, $\rho = 0.5\cdot10^{-4}$,
$N = 6\cdot10^{23}$, $\tau = 1$.

---

[11] [1] Cf. A. Einstein, *Zeitsch. f. Elektrochemie* 6 (1907).

[2] To be more precise, the square root of the mean value of Δ^2.

We obtain

$$\Delta = 0.8 \cdot 10^{-4} \text{ cm} = 0.8 \text{ micron}.$$

This number has an uncertainty of ± 25% due to the low accuracy with which N is known. [13]

It is of interest to compare the mean proper motion of the microscopic particles we just calculated with that of dissolved molecules or ions. For an undissociated dissolved substance whose coefficient of diffusion is known, Δ can be calculated from equation (7a). For sugar at room temperature we have $D = \frac{0.33}{24 \cdot 60 \cdot 60}$. From this we get from equation (7a) for $\tau = 1$ [14]

$$\Delta = 27.6 \text{ micron}.$$

From the number N and the molecular volume of solid sugar we can conclude that the diameter of a molecule of sugar is of the order of magnitude of a thousandth of a micron, i.e., about one thousand times smaller than the diameter of the suspended particle considered before. According to equation (8a), we can therefore expect Δ to be about $\sqrt{1000}$ times larger for sugar than for the particle with a diameter of 1 micron. As we have now seen, this is indeed approximately correct. [15]

For ions, Δ ["Δ" added by the translator] can be determined from their migration velocity ℓ from equation (8). ℓ equals the quantity of electricity in coulombs that would flow through 1 cm^2 per second at a concentration $\nu = 1$ of the ion in question and at a potential gradient of 1 volt per centimeter. In this imaginary process, the velocity v of the motion of the ions (in centimeter/second) is obviously given by the equation [16]

$$\ell = v \cdot 96{,}000.$$ [17]

Further, since 1 volt contains 10^8 electromagnetic units, and the charge of a (univalent) ion equals $\frac{9{,}600}{N}$ electromagnetic units, the force k exerted on one ion in the process imagined will be

$$k = \frac{10^8 \cdot 9{,}600}{N}.$$ [18]

Substituting this value of k and the value of v obtained from the previously found equation,

$$v = \frac{\ell}{96{,}000} ,$$

in equation (2), we get

$$\mathfrak{R} = \frac{k}{v} = \frac{10^8 \cdot 9{,}600 \cdot 96{,}000}{\ell \cdot N} .$$

With the usual definition of ℓ, this formula holds for polyvalent ions as well. Substituting this value for $\mathfrak{R}$ in equation (8), we get

$$\Delta = 4.25 \cdot 10^{-5} \sqrt{\ell T \tau} .$$

The formula yields for room temperature and $\tau = 1$

[19]

Ion	ℓ	Δ in microns
H	300	125
K	65	58
Diisoamylammonium ion $C_{10}H_{24}N$	24	35

Received on 1 April

Doc. 51

ON THE FUNDAMENTAL ELECTROMAGNETIC EQUATIONS FOR MOVING BODIES

by A. Einstein and J. Laub

[*Annalen der Physik* 26 (1908): 532-540]

In a recently published study[1] Mr. Minkowski has presented the fundamental equations for the electromagnetic processes in moving bodies. In view of the fact that this study makes rather great demands on the reader in its mathematical aspects, we do not consider it superfluous to derive here these important equations in an elementary way, which, is, by the way, essentially in agreement with that of Minkowski.

§*1. Derivation of the fundamental equations for moving bodies*

The route to be taken is as follows: We introduce two coordinate systems K and K', both of which are nonaccelerated but in relative motion. If the space contains matter at rest relative to K', then the laws of the electrodynamics of bodies at rest, described by the Maxwell-Hertz equations, will hold with respect to K'. If we transform these equations to the system K, we directly obtain the electrodynamic equations of moving bodies for the case that the velocity of the matter is spatially and temporally constant. Obviously, the equations so obtained hold at least in first approximation also in the case when the distribution of velocity of the matter is arbitrary. [2] This assumption is also partly justified by the fact that the result obtained in this way is strictly valid in the case of a number of bodies moving with different uniform velocities that are separated from each other by vacuum interspaces.

When referred to the system K', the vector of the electric force will be denoted by $\mathfrak{E}'$, that of the magnetic force by $\mathfrak{H}'$, that of the dielectric displacement by $\mathfrak{D}'$, that of the magnetic induction by $\mathfrak{B}'$, that of the

---

[1] H. Minkowski, *Göttinger Nachr.* 1908. [1]

electric current by $\mathfrak{s}'$; ρ' shall denote the electric density. For the reference system K' there shall hold the Maxwell-Hertz equations:

$$\text{curl}'\ \mathfrak{H}' = \frac{1}{c}\left[\frac{\partial \mathfrak{D}'}{\partial t'} + \mathfrak{s}'\right] , \tag{1}$$

$$\text{curl}'\ \mathfrak{E}' = -\frac{1}{c}\frac{\partial \mathfrak{B}'}{\partial t'} , \tag{2}$$

$$\text{div}'\ \mathfrak{D}' = \rho' , \tag{3}$$

$$\text{div}'\ \mathfrak{B}' = 0 . \tag{4}$$

Let us consider a second orthogonal reference system K whose axes are permanently parallel to those of K'. Let the origin of K' move in the positive direction of the x-axis of K with constant velocity v. As we know, according to the theory of relativity the following transformation equations[1] will then hold for every point event, provided the starting time point has been suitably chosen:

$$\left\{\begin{aligned} x' &= \beta(x - vt) \\ y' &= y , \\ z' &= z , \\ t' &= \beta\left[t - \frac{v}{c^2}x\right] , \end{aligned}\right. \qquad \left[\beta = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\right] ,$$

where x, y, z, t denote the space and time coordinates in the system K. If one carries out the transformations, one obtains the equations

$$\text{curl}\ \mathfrak{H} = \frac{1}{c}\left[\frac{\partial \mathfrak{D}}{\partial t} + \mathfrak{s}\right] , \tag{1a}$$

$$\text{curl}\ \mathfrak{E} = -\frac{1}{c}\frac{\partial \mathfrak{B}}{\partial t} , \tag{2a}$$

$$\text{div}\ \mathfrak{D} = \rho , \tag{3a}$$

---

[3] [1] A. Einstein, *Ann. d. Phys.* 17 (1905): 902.

$$\text{div}\,\mathfrak{B} = 0\ , \tag{4a}$$

where one has put

$$\left\{\begin{aligned}
\mathfrak{E}_x &= \mathfrak{E}'_x\ ,\\
\mathfrak{E}_y &= \beta\left[\mathfrak{E}'_y + \frac{v}{c}\,\mathfrak{B}'_z\right]\ ,\\
\mathfrak{E}_z &= \beta\left[\mathfrak{E}'_z - \frac{v}{c}\,\mathfrak{B}'_y\right]\ ,\\
\mathfrak{D}_x &= \mathfrak{D}'_x\ ,\\
\mathfrak{D}_y &= \beta\left[\mathfrak{D}'_y + \frac{v}{c}\,\mathfrak{H}'_z\right]\ ,\\
\mathfrak{D}_z &= \beta\left[\mathfrak{D}'_z - \frac{v}{c}\,\mathfrak{H}'_y\right]\ ,
\end{aligned}\right. \tag{6}$$

$$\left\{\begin{aligned}
\mathfrak{H}_x &= \mathfrak{H}'_x\ ,\\
\mathfrak{H}_y &= \beta\left[\mathfrak{H}'_y - \frac{v}{c}\,\mathfrak{D}'_z\right]\ ,\\
\mathfrak{H}_z &= \beta\left[\mathfrak{H}'_z + \frac{v}{c}\,\mathfrak{D}'_y\right]\ ,\\
\mathfrak{B}_x &= \mathfrak{B}'_x\ ,\\
\mathfrak{B}_y &= \beta\left[\mathfrak{B}'_y - \frac{v}{c}\,\mathfrak{E}'_z\right]\ ,\\
\mathfrak{B}_z &= \beta\left[\mathfrak{B}'_z + \frac{v}{c}\,\mathfrak{E}'_y\right]\ ,
\end{aligned}\right. \tag{7}$$

and

$$\rho = \beta\left[\rho' + \frac{v}{c}\,\mathfrak{s}'_x\right]\ , \tag{8}$$ [4]

$$\left\{\begin{aligned}
\mathfrak{s}_x &= \beta\left[\mathfrak{s}'_x + \frac{v}{c}\,\rho'\right]\ ,\\
\mathfrak{s}_y &= \mathfrak{s}'_y\ ,\\
\mathfrak{s}_z &= \mathfrak{s}'_z\ .
\end{aligned}\right. \tag{9}$$ [5]

To obtain the expressions for the primed quantities as functions of the unprimed ones, one has to interchange the primed and unprimed quantities and to replace v by $-v$.

The equations (1a) to (4a), which describe the electromagnetic processes relative to the system K, have the same form as the equations (1) to (4). *We will therefore* use the same terminology *for the quantities*

$$\mathfrak{E},\ \mathfrak{D},\ \mathfrak{H},\ \mathfrak{B},\ \rho,\ \mathfrak{s}$$

as for the corresponding quantities relative to the system K'. *Thus* $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$, ρ, $\mathfrak{s}$ *are the electric force, the dielectric displacement, the magnetic force, the magnetic induction, the electric density, the electric current* with respect to K.

For vacuum the transformation equations (6) and (7) reduce to the equations for electric and magnetic forces found earlier.[1]

It is clear that by repeated application of transformations of the kind that we have just performed one must always arrive at equations of the same form as the original equations (1) to (4), and that for such transformations equations (6) to (9) apply, since formally the transformation did not make use of the fact that the matter was at rest relative to the original system K'.

We assume that the transformed equations (1a) to (4a) are also valid if the velocity of the matter is spatially and temporally variable, which will be [6] correct in the first approximation.

It is remarkable that the boundary conditions for the vectors $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$ at the boundary of two media are the same as for bodies at rest. This [7] follows directly from equations (1a) to (4a).

Just like equations (1) to (4), equations (1a) to (4a) hold quite generally for inhomogeneous and anisotropic bodies. They do not completely determine the electromagnetic processes, however. Rather, relations that express the vectors $\mathfrak{D}$, $\mathfrak{B}$, and $\mathfrak{s}$ as functions of $\mathfrak{E}$ and $\mathfrak{H}$ need to be given in addition. We will now give such equations for the case that the *matter is isotropic*. If we first consider the case when all matter is at rest relative to K', then the following equations hold with respect to K':

---

[1] A. Einstein, *loc. cit.*, p. 909.

$$\mathfrak{D}' = \epsilon\mathfrak{E}' \,, \tag{10}$$

$$\mathfrak{B}' = \mu\mathfrak{H}' \,, \tag{11}$$

$$\mathfrak{s}' = \sigma\mathfrak{E}' \,, \tag{12}$$

where ϵ = dielectric constant, μ = permeability, and σ = electrical conductivity are to be regarded as known functions of x', y', z', t'. By transforming (10) to (12) to K by means of inversion of our transformation equations (6) to (9), one obtains the relations holding for the system K:

$$\left\{\begin{aligned} &\mathfrak{D}_x = \epsilon\, \mathfrak{E}_x \,, \\ &\mathfrak{D}_y - \frac{v}{c}\mathfrak{H}_z = \epsilon\left[\mathfrak{E}_y - \frac{v}{c}\mathfrak{B}_z\right] , \\ &\mathfrak{D}_z + \frac{v}{c}\mathfrak{H}_y = \epsilon\left[\mathfrak{E}_z + \frac{v}{c}\mathfrak{B}_y\right] , \end{aligned}\right. \tag{10a}$$

$$\left\{\begin{aligned} &\mathfrak{B}_x = \mu\, \mathfrak{H}_x \,, \\ &\mathfrak{B}_y + \frac{v}{c}\mathfrak{E}_z = \mu\left[\mathfrak{H}_y + \frac{v}{c}\mathfrak{D}_z\right] , \\ &\mathfrak{B}_z - \frac{v}{c}\mathfrak{E}_y = \mu\left[\mathfrak{H}_z - \frac{v}{c}\mathfrak{D}_y\right] , \end{aligned}\right. \tag{11a}$$

$$\left\{\begin{aligned} &\beta\left[\mathfrak{s}_x - \frac{v}{c}\rho\right] = \sigma\, \mathfrak{E}_x \,, \\ &\mathfrak{s}_y = \sigma\beta\left[\mathfrak{E}_y - \frac{v}{c}\mathfrak{B}_z\right] , \\ &\mathfrak{s}_z = \sigma\beta\left[\mathfrak{E}_z + \frac{v}{c}\mathfrak{B}_y\right] . \end{aligned}\right. \tag{12a}$$
[8]

If the velocity of the matter is not parallel to the X-axis, but is determined by the vector $\mathfrak{v}$ instead, one obtains the vectorial relations that are analogous to equations (10a) to (12a):

[9]

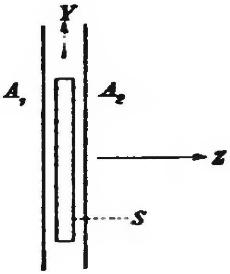
$$(13)\quad \left\{\begin{aligned} \mathfrak{D} + \frac{1}{c}[\mathfrak{v}\mathfrak{H}] &= \epsilon\left\{\mathfrak{E} + \frac{1}{c}[\mathfrak{v}\mathfrak{B}]\right\}, \\ \mathfrak{B} - \frac{1}{c}[\mathfrak{v}\mathfrak{E}] &= \mu\left\{\mathfrak{H} - \frac{1}{c}[\mathfrak{v}\mathfrak{D}]\right\}, \\ \beta\left[\mathfrak{s}_{\mathfrak{v}} - \frac{|\mathfrak{v}|}{c}\rho\right] &= \sigma\left\{\mathfrak{E} + \frac{1}{c}[\mathfrak{v}\mathfrak{B}]\right\}_{\mathfrak{v}}, \\ \mathfrak{s}_{\bar{\mathfrak{v}}} &= \sigma\beta\left\{\mathfrak{E} + \frac{1}{c}[\mathfrak{v}\mathfrak{B}]\right\}_{\bar{\mathfrak{v}}}, \end{aligned}\right.$$

where the subscript $\mathfrak{v}$ signifies that the component has to be taken in the direction of $\mathfrak{v}$, and the subscript $\bar{\mathfrak{v}}$ that the components have to be taken in the directions $\bar{\mathfrak{v}}$ perpendicular to $\mathfrak{v}$.

§2. *On the electromagnetic behavior of moving dielectrics. Wilson's experiment*

In the following section we will use a simple special case to show how moving dielectrics behave according to the theory of relativity and how the results differ from those obtained by the Lorentz theory.

Let S be a prismatic strip of a homogeneous, isotropic nonconductor, indicated in its cross section (cf. figure), that extends to infinity perpendicularly to the plane of the paper in both directions and that moves with the



constant velocity v away from the observer toward the plane of the paper between the condenser plates A_1 and A_2. The dimension of the strip S perpendicular to the plates A shall be infinitesimally small compared to its dimension parallel to the plates and to the dimensions of the plates A; also, the gap between S and the plates A (henceforth called gap in brief) shall be negligible compared to the thickness of S. We refer the system of body under consideration to a coordinate system that is at rest relative to the plates A and whose positive X-direction shall coincide with the direction of motion, while its Y- and Z-axes, respectively, are parallel and perpendicular to the plates A. We will examine the electromagnetic behavior of the piece of the strip located between the plates A if the electromagnetic state is stationary.

We imagine a closed surface that just encloses the effective part of the condenser plates together with that of the piece of the strip lying between them. Since no moving true charges nor electric conduction currents exist within this surface, the equations (cf. equations (1a) to (4a))

$$\operatorname{curl} \mathfrak{H} = 0 \,,$$
$$\operatorname{curl} \mathfrak{E} = 0 \,,$$

apply. Thus, within this space the electric as well as the magnetic force is derivable from a potential. Hence we immediately know the distribution of the vectors $\mathfrak{E}$ and $\mathfrak{H}$ if the distribution of the free electric and magnetic density, respectively, is known. We shall limit ourselves to consideration of the case in which the magnetic force $\mathfrak{H}$ is parallel to the Y-axis, and the electric force $\mathfrak{E}$ is parallel to the Z-axis. We are justified in doing this, as well as in assuming that the pertinent fields both within the strip and within the gap are homogeneous, due to the conditions stipulated earlier regarding the orders of magnitudes of the dimensions of the system under consideration. We also conclude immediately that the magnetic masses at the ends of the strip cross section make only a vanishingly small contribution to the magnetic field.[1] Equations (13) then yield the following relations for the interior of the strip:

---

[1] This is also evident from the fact that, without essential change of the conditions, we could give the condenser plates and the strip a circular cylindric shape, in which case, for reasons of symmetry, free magnetic masses could not arise at all.

$$\mathfrak{D}_z + \frac{v}{c}\,\mathfrak{H}_y = \epsilon\left[\mathfrak{E}_z + \frac{v}{c}\,\mathfrak{B}_y\right] ,$$

$$\mathfrak{B}_y + \frac{v}{c}\,\mathfrak{E}_z = \mu\left[\mathfrak{H}_y + \frac{v}{c}\,\mathfrak{D}_z\right] .$$

These equations can also be written in the following form:

$$(1)\qquad \left\{\begin{aligned} &\left[1 - \epsilon\mu\,\frac{v^2}{c^2}\right]\mathfrak{B}_y = \frac{v}{c}(\epsilon\mu - 1)\mathfrak{E}_z + \mu\left[1 - \frac{v^2}{c^2}\right]\mathfrak{H}_y ,\\ &\left[1 - \epsilon\mu\,\frac{v^2}{c^2}\right]\mathfrak{D}_z = \epsilon\left[1 - \frac{v^2}{c^2}\right]\mathfrak{E}_z + \frac{v}{c}(\epsilon\mu - 1)\mathfrak{H}_y .\end{aligned}\right.$$

Concerning the interpretation of (1) we remark the following: The dielectric displacement $\mathfrak{D}_z$ experiences no jump at the surface of the strip, hence it equals the charge of the condenser plates (more exactly, of the plate A_1) per unit area. Further, $\mathfrak{E}_z \times \delta$ equals the potential difference between the condenser plates A_1 and A_2 if δ denotes the separation of the plates, because if one imagines that the strip is separated by an infinitely narrow slit running parallel to the XZ-plane, then $\mathfrak{E}$ equals the electric force in the slit on account of the boundary conditions holding for that vector.

Next we consider the case that no magnetic field excited from the outside is present, i.e., according to the above, that in the space considered the magnetic field strength vanishes. Then equations (1) will have the following form:

$$\left[1 - \epsilon\mu\,\frac{v^2}{c^2}\right]\mathfrak{B}_y = \frac{v}{c}(\epsilon\mu - 1)\mathfrak{E}_z ,$$

$$\left[1 - \epsilon\mu\,\frac{v^2}{c^2}\right]\mathfrak{D}_z = \epsilon\left[1 - \frac{v^2}{c^2}\right]\mathfrak{E}_z .$$

Since we must have $v < c$, the coefficients of $\mathfrak{E}_z$ in the last two equations must be positive if $\epsilon\mu - 1 > 0$. In contrast, the coefficients of $\mathfrak{B}_y$ and $\mathfrak{D}_z$ are larger, equal to, or smaller than zero, respectively, depending on whether the velocity of the strip is smaller, equal to, or larger than $c/\sqrt{\epsilon\mu}$, i.e., than the velocity of the electromagnetic waves in the strip medium. Hence, if $\mathfrak{E}_z$ has a fixed value, i.e., if one applies a fixed potential difference to the condenser plates, and varies the strip velocity from lower to higher values, then at first both the charge of the condenser plates, which is proportional to the vector $\mathfrak{D}$, and the magnetic induction $\mathfrak{B}$ in the strip will increase. When v reaches the value $c/\sqrt{\epsilon\mu}$, both the condenser's charge and the magnetic induction become infinitely large. Hence, in this case even an arbitrarily small applied potential difference would destroy the strip. For all $v > c/\sqrt{\epsilon\mu}$ there result negative values for $\mathfrak{D}$ and $\mathfrak{B}$. Thus, in the last case a potential difference applied to the condenser plates would charge the condenser in the sense opposite to the potential difference.

Finally, we consider the case of the presence of a magnetic field $\mathfrak{H}_y$ excited from the outside. We then have the equation

$$\left[1 - \epsilon\mu \frac{v^2}{c^2}\right]\mathfrak{D}_z = \epsilon\left[1 - \frac{v^2}{c^2}\right]\mathfrak{E}_z + \frac{v}{c}(\epsilon\mu - 1)\mathfrak{H}_y ,$$

which yields a relation between $\mathfrak{E}_z$ and $\mathfrak{D}_z$ at a given $\mathfrak{H}_y$. If one restricts oneself to quantities of the first order in v/c, one has

$$\mathfrak{D}_z = \epsilon\mathfrak{E}_z + \frac{v}{c}(\epsilon\mu - 1)\mathfrak{H}_y , \tag{2}$$

while Lorentz's theory leads to the expression

$$\mathfrak{D}_z = \epsilon\mathfrak{E}_z + \frac{v}{c}(\epsilon - 1)\mu\mathfrak{H}_y . \tag{3}$$ [11]

As we know, the latter equation has been experimentally tested by H. A. Wilson (Wilson effect). One sees that (2) and (3) differ in terms of first [12] order. If we would have a dielectric body of considerable permeability, it would be possible to decide experimentally between equations (2) and (3). [13]

If one connects the plates A_1 and A_2 by a conductor, a charge of magnitude $\mathfrak{D}_z$ per unit area is generated on the condenser plates; one obtains it from equation (2) by taking into account that for connected condenser plates, $\mathfrak{E}_z = 0$. One gets

$$\mathfrak{D}_z = \frac{v}{c}(\epsilon\mu - 1)\mathfrak{H}_y .$$

If one connects the condenser plates A_1 and A_2 with an electrometer of infinitesimally small capacity, then $\mathfrak{D}_z = 0$, and one obtains for the potential difference $(\mathfrak{E}_z . \delta)$ the equation

$$0 = \epsilon \mathfrak{E}_z + \frac{v}{c}(\epsilon\mu - 1)\mathfrak{H}_y \ .$$

Bern, 29 April 1908. (Received on 2 May 1908)

Doc. 52

ON THE PONDEROMOTIVE FORCES EXERTED ON BODIES AT REST IN THE ELECTROMAGNETIC FIELD

by A. Einstein and J. Laub

[*Annalen der Physik* 26 (1908): 541-550]

In a recently published study[1] Mr. Minkowski presented an expression for the ponderomotive forces of electromagnetic origin that act on arbitrarily moving bodies. If one specializes Minkowski's expressions to isotropic, [2] homogeneous bodies at rest, one obtains for the X-component of the force acting on a unit volume

$$K_x = \rho\mathfrak{E}_x + \mathfrak{s}_y\mathfrak{B}_z - \mathfrak{s}_z\mathfrak{B}_y , \tag{1}$$ [3]

where ρ denotes the electric density, $\mathfrak{s}$ the electric conduction current, $\mathfrak{E}$ the electric field strength, and $\mathfrak{B}$ the magnetic induction. This expression seems to us not to be in agreement with the electron-theoretical picture for the following reasons: While a body traversed by an electric current (conduction current) experiences a force in the magnetic field, according to equation (1) this would not be the case if the body in the magnetic field were permeated by a polarization current $(\partial\mathfrak{D}/\partial t)$ instead of a conduction current. Thus, according to Minkowski there exists here a difference in principle between a displacement current and a conduction current such that a conductor cannot be considered as a dielectric with an infinitely large dielectric constant.

In view of this state of affairs, it seems to us that it would be of interest to derive the ponderomotive forces for arbitrary magnetizable bodies on the basis of the electron theory. We present here such a derivation, though restricting ourselves to bodies at rest.

---

[1] H. Minkowski, *Gött. Nachr.* (1908), p. 45. [1]

§1. *Forces that do not depend on the velocities of elementary particles*

In this derivation we will consistently base ourselves on the standpoint [4] of the electron theory[1]; hence we put

$$\mathfrak{D} = \mathfrak{E} + \mathfrak{P}, \tag{2}$$

$$\mathfrak{B} = \mathfrak{H} + \mathfrak{Q}, \tag{3}$$

where $\mathfrak{P}$ denotes the electric and $\mathfrak{Q}$ the magnetic polarization vector. We think of electric and magnetic polarizations, respectively, as consisting of spatial displacements of electric and magnetic mass particles of dipoles that are bound to equilibrium positions. In addition, we also assume the presence of mobile electric particles not bound to dipoles (conduction electrons). Let Maxwell's equations for empty space be valid in the space between the above particles, and let, as in Lorentz, the *interactions between matter and electromagnetic field be exclusively brought about by these particles.* [6] Accordingly, we assume that the forces exerted by the electromagnetic field on the volume element of the matter equal the resultant of the ponderomotive forces exerted by this field on all elementary electric and magnetic particles in the volume element considered. By a volume element of the matter we always understand a space so large that it contains a very large number of electric and magnetic particles. The boundaries of a volume element must always be imagined as drawn such that the boundary surface does not cut through any electric or magnetic dipoles.

First we calculate that force acting on a dipole which is due to the field strength $\mathfrak{E}$ not being exactly the same at the locations of the elementary masses of the dipole. If $\mathfrak{p}$ denotes the vector of the dipole moment, one obtains the following expression for the X-component of the force sought:

$$\mathfrak{f}_x = \mathfrak{p}_x \frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{p}_y \frac{\partial \mathfrak{E}_x}{\partial y} + \mathfrak{p}_z \frac{\partial \mathfrak{E}_x}{\partial z} .$$

---

[5] [1] However, we stick to the dual treatment of electric and magnetic phenomena for the sake of a simpler presentation.

If one imagines that the above expression is formed for and summed over all dipoles in the unit volume, one obtains, taking into account the relation

$$\sum \mathfrak{p} = \mathfrak{P} \,,$$

the equation

$$\mathfrak{F}_{1x} = \left\{\mathfrak{P}_x \frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{P}_y \frac{\partial \mathfrak{E}_x}{\partial y} + \mathfrak{P}_z \frac{\partial \mathfrak{E}_x}{\partial z}\right\} . \tag{4}$$

If the algebraic sum of the positive and negative conduction electrons does not vanish, then the expression (4) contains an additional term, which we shall now calculate. The X-component of the ponderomotive force acting on a conduction electron of electric mass e is $e\mathfrak{E}_x$. If one sums over all conduction electrons of the unit volume, one obtains

$$\mathfrak{F}_{2x} = \mathfrak{E}_x \sum e \,. \tag{5}$$

If one imagines that the matter in the unit volume is enclosed by a surface that does not cut through any dipole, one obtains in accordance with Gauss's law and the definition of the displacement vector $\mathfrak{D}$

$$\sum e = \operatorname{div} \mathfrak{D} \,,$$

so that

$$\mathfrak{F}_{2x} = \mathfrak{E}_x \operatorname{div} \mathfrak{D} \,. \tag{5a}$$

The X-component of the force exerted by the electric field strength on the unit volume of the matter therefore equals

$$\mathfrak{F}_{e_x} = \mathfrak{F}_{1x} + \mathfrak{F}_{2x} = \mathfrak{P}_x \frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{P}_y \frac{\partial \mathfrak{E}_x}{\partial y} + \mathfrak{P}_z \frac{\partial \mathfrak{E}_x}{\partial z} + \mathfrak{E}_x \operatorname{div} \mathfrak{D} \,. \tag{6}$$

In an analogous way, taking into account the relation

$$\operatorname{div} \mathfrak{B} = 0,$$

we obtain for the X-component of the force imparted by the magnetic field strength

$$(7) \qquad \mathfrak{F}_{mx} = \left\{\mathfrak{Q}_x \frac{\partial \mathfrak{H}_x}{\partial x} + \mathfrak{Q}_y \frac{\partial \mathfrak{H}_x}{\partial y} + \mathfrak{Q}_z \frac{\partial \mathfrak{H}_x}{\partial z}\right\}$$

It should be noted that the derivation of the expressions (6) and (7) does not require any assumptions about the relation connecting the field strengths $\mathfrak{E}$ and $\mathfrak{H}$ with the polarization vectors $\mathfrak{P}$ and $\mathfrak{Q}$.

In the case of anisotropic bodies, the electric and magnetic field strengths, respectively, do not impart only a force, but also force couples that act on the matter. The torque sought can easily be obtained for the individual dipoles and summation over all electric and magnetic dipoles in the unit volume. One obtains

$$(8) \qquad \mathfrak{L} = \{[\mathfrak{P}\mathfrak{E}] + [\mathfrak{Q}\mathfrak{H}]\} .$$

Formula (6) yields those ponderomotive forces that play a role in electrostatic problems. We want to transform this equation, applied in the case of isotropic bodies, in such a way that it allows a comparison with the expression for ponderomotive forces used in electrostatics. If we put

$$\mathfrak{P} = (\epsilon - 1)\mathfrak{E} ,$$

equation (6) becomes

$$\mathfrak{F}_{e_x} = \mathfrak{E}_x \operatorname{div} \mathfrak{D} - \frac{1}{2}\mathfrak{E}^2 \frac{\partial \epsilon}{\partial x} + \frac{1}{2}\frac{\partial}{\partial x}(\epsilon - 1)\mathfrak{E}^2 .$$

The first two terms of this expression are identical with those familiar from electrostatics. As one can see, the third term is derivable from a potential. If the forces involved act upon a body in the vacuum, this term does not contribute anything on integration over the body. However, if the ponderomotive forces involved act on liquids, then the part of the force corresponding to the third term is compensated by a pressure distribution in the liquid when in equilibrium.

§2. *Forces that depend on the velocities of the elementary particles*

We now turn to the part of the ponderomotive force that is produced by the velocities of motion of the elementary charges.

We start from the Biot-Savart law. According to experience, the force that acts on a unit volume of a volume element traversed by a current and located in a magnetic field is

$$\frac{1}{c}\,[\mathfrak{s}\mathfrak{H}]\ ,$$

if the matter traversed by the current is not magnetically polarizable. As far as we know, for the interior of a magnetically polarizable body that force has so far been set equal[1] to

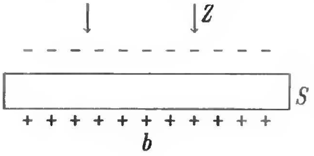
$$\frac{1}{c}\,[\mathfrak{s}\mathfrak{B}]\ ,$$

where $\mathfrak{B}$ denotes the magnetic induction. We will now show that the force acting on the current-carrying volume element is *also* obtained in the case where the current-traversed matter is *magnetically polarizable* if the volume force

(9) $$\mathfrak{F}_{\mathfrak{s}} = \frac{1}{c}\,[\mathfrak{s}\mathfrak{H}]$$ [8]

is added to the force expressed by equation (7). We will first illustrate this by a simple example.

Let the infinitely thin strip *S*, plotted in its cross section, stretch to infinity in both directions perpendicular to the plane of the paper. Assume that it consists of magnetically polarizable material and is located in a homogeneous magnetic field $\mathfrak{H}_a$, whose direction is indicted by the arrows (cf. figure). We ask for the force acting on the material strip if the latter is traversed by a current i.



---

[1] Cf., e.g.,also M. Abraham, *Theorie der Elektrizität* 2 (1905): 319. [7]

Experience shows that this force is independent of the magnetic permeability of the material of the conductor, and from this it has been concluded that it is not the field strength $\mathfrak{H}$, but the magnetic induction $\mathfrak{B}_i$, that is relevant for the ponderomotive force, because in the interior of the strip the magnetic induction $\mathfrak{B}_i$ is equal to the force $\mathfrak{H}_a$ acting outside the strip independent of the value of the strip's permeability, while for a given external field the force in the interior of the strip depends on μ. However, this argument is not conclusive, because the ponderomotive force considered is not the only one acting on our material strip, since the external field $\mathfrak{H}_a$ induces magnetic layers of density[1] $\mathfrak{H}_a(1 - 1/\mu)$ on the top and bottom sides of the material strip, the layer being negative on the top, and positive on the bottom. Each of these layers is acted on by a force which is produced by the current flowing through the strip; this magnetic force has a strength of $\mathrm{i}/2b$ per unit length of the strip[2] and its direction on the top differs from that on the bottom. The ponderomotive forces so obtained add up, so that we get the ponderomotive force $(1 - 1/\mu)\mathfrak{H}_a \mathrm{i}$. It seems that this force has not been taken into consideration up to now.

The total force exerted per unit length of our strip is, then, equal to the sum of the force just calculated and the force R exerted by the magnetic field on the strip's volume elements due to the current flow. Since experience shows that the total ponderomotive force acting on the unit length is $\mathrm{i}\mathfrak{H}_a$, we have the equation

$$(1 - \frac{1}{\mu})\mathrm{i}\mathfrak{H}_a + R = \mathrm{i}\mathfrak{H}_a$$

or

$$R = \frac{\mathrm{i}\mathfrak{H}_a}{\mu} = \mathrm{i}\mathfrak{H}_i \ .$$

This shows that for the calculation of the ponderomotive force R, which is acting on the current-traversed volume elements, it is the field strength $\mathfrak{H}_i$ and not the induction $\mathfrak{B}_i$ that is responsible.

---

[1] Since the density is

$$\mathfrak{Q}_i = \mathfrak{B}_i - \mathfrak{H}_i = \mathfrak{H}_a(1 - \frac{1}{\mu}) \ .$$

[2] Strictly speaking, based on the results of the previous section, we should have introduced volume forces instead of the above forces acting on the surface layers, but this is of no consequence.

To eliminate all doubt, we will discuss one more example which shows that the principle of equality of action and reaction requires the *Ansatz* we have chosen.

We envision a cylindric conductor surrounded by empty space and traversed by the current $\mathfrak{s}$, which stretches to infinity along the X-axis of a coordinate system in both directions. The material constants of the conductor, as well as the field vectors considered in the following, shall be independent of x, but shall be functions of y and z. The conductor shall be a magnetically hard body and shall have a magnetization perpendicular to the X-axis. We assume that no external field acts on the conductor, and thus the magnetic force $\mathfrak{H}$ vanishes far from the conductor.

It is clear that no ponderomotive force acts on the conductor as a whole, because no reaction opposing such an action can be specified. We now want to show that the above force indeed vanishes given the *Ansatz* we have chosen. In accordance with equations (7) and (9), the entire force acting on the unit length of our conductor in the direction of the Z-axis can be presented in the form

$$R = \int\left[\Omega_y \frac{\partial \mathfrak{H}_z}{\partial y} + \Omega_z \frac{\partial \mathfrak{H}_z}{\partial z}\right] df + \int \frac{1}{c}\, \mathfrak{s}_x \mathfrak{H}_y df \ , \tag{10}$$

where df denotes a surface element of the YZ-plane. We assume that all pertinent quantities are continuous on the surface of the conductor. First we consider the first integral of equation (10). We have

$$\Omega_y \frac{\partial \mathfrak{H}_z}{\partial y} + \Omega_z \frac{\partial \mathfrak{H}_z}{\partial z} = \frac{\partial \Omega_y \mathfrak{H}_z}{\partial y} + \frac{\partial \Omega_z \mathfrak{H}_z}{\partial z} - \mathfrak{H}_z\left[\frac{\partial \Omega_y}{\partial y} + \frac{\partial \Omega_z}{\partial z}\right] .$$

If one substitutes the right-hand side of this equation in our integral, then the first two terms vanish on integration over the YZ-plane because the forces vanish at infinity. Taking into consideration that

$$\operatorname{div} \mathfrak{B} = 0 \ ,$$

the third term can be transformed so that our integral assumes the form

$$\int \mathfrak{H}_z\left[\frac{\partial \mathfrak{H}_y}{\partial y} + \frac{\partial \mathfrak{H}_z}{\partial z}\right] df \; .$$

Now we have

$$\mathfrak{H}_z\left[\frac{\partial \mathfrak{H}_y}{\partial y} + \frac{\partial \mathfrak{H}_z}{\partial z}\right] = \frac{\partial \mathfrak{H}_y \mathfrak{H}_z}{\partial y} + \frac{1}{2}\,\frac{\partial \mathfrak{H}_z^2}{\partial z} - \mathfrak{H}_y\,\frac{\partial \mathfrak{H}_z}{\partial y} \; .$$

However, the two terms $\frac{\partial \mathfrak{H}_y \mathfrak{H}_z}{\partial y} + \frac{1}{2}\,\frac{\partial \mathfrak{H}_z^2}{\partial z}$ vanish on integration. Using Maxwell's equations, the term $-\mathfrak{H}_y\,\frac{\partial \mathfrak{H}_z}{\partial y}$ can be transformed to

$$-\frac{1}{c}\,\mathfrak{H}_y\left\{\mathfrak{s}_x + \frac{\partial \mathfrak{H}_y}{\partial z}\right\} ,$$

so that finally we can write equation (10) as

$$R = -\frac{1}{c}\int \mathfrak{H}_y\left\{\mathfrak{s}_x + \frac{\partial \mathfrak{H}_y}{\partial z}\right\} df + \frac{1}{c}\int \mathfrak{s}_x \mathfrak{H}_y\, df$$

$$= -\frac{1}{c}\int \mathfrak{H}_y\,\frac{\partial \mathfrak{H}_y}{\partial z}\, df = -\frac{1}{2c}\int \frac{\partial \mathfrak{H}_y^2}{\partial z}\, df \; .$$

The last integral equals zero, because the forces vanish at infinity. –

Thus having ascertained the force that acts on matter traversed by a conduction current, we obtain the force that acts on a body permeated by a polarization current by noting that from the standpoint of the theory of electrons the polarization current and the conduction current are completely equivalent with regard to electrodynamic action.

By taking into account the duality of magnetic and electric phenomena, one also obtains the force exerted on a body permeated by a magnetic polarization current in the electric field. In this way we obtain the following equations as an overall expression for the forces that depend on the velocity of the elementary particles:

$$(11) \qquad \mathfrak{F}_a = \frac{1}{c}[\mathfrak{s}\mathfrak{H}] + \frac{1}{c}\left[\frac{\partial \mathfrak{P}}{\partial t}\,\mathfrak{H}\right] + \frac{1}{c}\left[\mathfrak{E}\,\frac{\partial \Omega}{\partial t}\right] .$$

§3. *Equality of action and reaction*

Adding equations (6), (7), and (11), one obtains the overall expression for the *X*-component of the ponderomotive force acting on the matter per unit volume in the form

$$\mathfrak{F}_x = \mathfrak{E}_x \operatorname{div} \mathfrak{D} + \mathfrak{P}_x \frac{\partial \mathfrak{E}_x}{\partial x} + \mathfrak{P}_y \frac{\partial \mathfrak{E}_x}{\partial y} + \mathfrak{P}_z \frac{\partial \mathfrak{E}_x}{\partial z}$$

$$+ \mathfrak{M}_x \frac{\partial \mathfrak{H}_x}{\partial x} + \mathfrak{M}_y \frac{\partial \mathfrak{H}_x}{\partial y} + \mathfrak{M}_z \frac{\partial \mathfrak{H}_x}{\partial z}$$

$$+ \frac{1}{c}[\mathfrak{s}\mathfrak{H}]_x + \frac{1}{c}\left[\frac{\partial \mathfrak{P}}{\partial t}\, \mathfrak{H}\right]_x + \frac{1}{c}\left[\mathfrak{E}\, \frac{\partial \mathfrak{M}}{\partial t}\right]_x .$$

This equation can also be written as

$$\mathfrak{F}_x = \mathfrak{E}_x \operatorname{div} \mathfrak{E} + \frac{1}{c}[\mathfrak{s}\mathfrak{H}]_x + \frac{1}{c}\left[\frac{\partial \mathfrak{D}}{\partial t}\, \mathfrak{H}\right]_x + \mathfrak{H}_x \operatorname{div} \mathfrak{H} + \frac{1}{c}\left[\mathfrak{E}\, \frac{\partial \mathfrak{B}}{\partial t}\right]_x$$

$$+ \frac{\partial(\mathfrak{P}_x \mathfrak{E}_x)}{\partial x} + \frac{\partial(\mathfrak{P}_y \mathfrak{E}_x)}{\partial y} + \frac{\partial(\mathfrak{P}_z \mathfrak{E}_x)}{\partial z}$$

$$+ \frac{\partial(\mathfrak{M}_x \mathfrak{H}_x)}{\partial x} + \frac{\partial(\mathfrak{M}_y \mathfrak{H}_x)}{\partial y} + \frac{\partial(\mathfrak{M}_z \mathfrak{H}_x)}{\partial z} - \frac{1}{c}\,\frac{\partial}{\partial t}[\mathfrak{E}\mathfrak{H}]_x .$$

If one makes use of Maxwell's equations to replace

$$\frac{1}{c}\left[\sigma + \frac{\partial \mathfrak{D}}{\partial t}\right] \quad \text{and} \quad \frac{1}{c}\,\frac{\partial \mathfrak{B}}{\partial t}$$

by curl $\mathfrak{H}$ and curl $\mathfrak{E}$, respectively, a simple transformation yields [9]

$$\mathfrak{F}_x = \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z} - \frac{1}{c^2}\,\frac{\partial \mathfrak{s}_z}{\partial t} , \tag{12}$$

[10] where we have put[1]

$$(13)\quad \begin{cases} X_x = -\frac{1}{2}(\mathfrak{E}^2 + \mathfrak{H}^2) + \mathfrak{E}_x\mathfrak{D}_x + \mathfrak{H}_x\mathfrak{B}_x, \\ X_y = \mathfrak{E}_x\mathfrak{D}_y + \mathfrak{H}_x\mathfrak{B}_y, \\ X_z = \mathfrak{E}_x\mathfrak{D}_z + \mathfrak{H}_x\mathfrak{B}_z, \\ \mathfrak{S}_x = c[\mathfrak{E}\mathfrak{H}]_x . \end{cases}$$

Corresponding equations hold for the other two components of the ponderomotive force.

By integrating (12) over the infinite space, one obtains the equation

[12] $$(14)\quad \int \mathfrak{F}_x d\tau = -\frac{1}{c^2}\int d\tau \frac{d\mathfrak{S}_x}{dt} ,$$

if the field vectors vanish at infinity. This equations states that on introduction of the electromagnetic momentum our ponderomotive forces satisfy the law of equality of action and reaction.

Bern, 7 May 1908. (Received on 13 May 1908)

---

[1] Geheimrat Wien kindly drew our attention to the fact that H. A. Lorentz had already presented the ponderomotive forces for nonmagnetizable bodies in this form. *Enzyklopädie der mathematischen Wissenschaften* 5, p.247. [11]

Doc. 53

CORRECTION TO THE PAPER: "ON THE FUNDAMENTAL ELECTROMAGNETIC EQUATIONS FOR MOVING BODIES"

by A. Einstein and J. Laub

[*Annalen der Physik* 27 (1908): 232]

Two errors had slipped into the paper published under the above title in this journal (26, p. 532, 1908): [1]

p. 534, formula (8) should read

$$\rho = \beta(\rho' + \frac{v}{c^2}\mathfrak{s}'_x)$$

instead of

$$\rho = \beta(\rho' + \frac{v}{c}\mathfrak{s}'_x) \ ,$$

and the first of the formulas (9),

$$\mathfrak{s}_x = \beta(\mathfrak{s}'_x + v\rho')$$

instead of

$$\mathfrak{s}_x = \beta(\mathfrak{s}'_x + \frac{v}{c}\rho').$$

Likewise, the first of formulas (12a) and the third of formulas (13) should read

$$\beta(\mathfrak{s}_x - v\rho) = \sigma\mathfrak{E}_x$$

and

$$\beta(\mathfrak{s}'_{\mathfrak{v}} - |\mathfrak{v}|\rho) = \sigma\left\{\mathfrak{E} + \frac{1}{c}[\mathfrak{v}\mathfrak{B}]\right\}_{\mathfrak{v}}.$$

(Received on 24 August 1908)

Doc. 54

REMARKS ON OUR PAPER: "ON THE FUNDAMENTAL ELECTROMAGNETIC EQUATIONS FOR MOVING BODIES"

by A. Einstein and J. Laub

[*Annalen der Physik*, 28 (1908): 445-447]

Mr. Laue was kind enough to call our attention to an inaccuracy in our
[2] paper cited in the title.[1] We say there (Ann. d. Phys. *26*(1908): 535):

"It is remarkable that the boundary conditions for the vectors $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$ at the boundary of two media are the same as for bodies at rest. This follows directly from equations (1a) to (4a)."

Apart from the fact that equations (3a) and (4a) are irrelevant for the derivation of the boundary conditions, this statement is correct only if the component of the motion normal to the boundary surface vanishes, which is actually the case in the problem treated in §2 of the paper quoted. The boundary conditions of the general case are most easily found in the following
[3] way, which corresponds to the one taken by Heinrich Hertz.

If the boundary surface, or, more exactly, the infinitely thin boundary transition shell, moves in some arbitrary manner, then, in a point situated in it and instantaneously at rest, the quantities determining the electromagnetic field will in general vary discontinuously, infinitely fast, with time; however, these changes will be continuous for a point *moving with the matter*. Thus, the application of the operator

$$\frac{\partial}{\partial t} + (\mathfrak{v}\nabla)$$

to a scalar or to a vector will not lead to infinitely large values even in the boundary surface. If we then write equation (1a)[2] in the form

$$\frac{1}{c}\left[\frac{\partial\mathfrak{D}}{\partial t} + (\mathfrak{v}\nabla)\mathfrak{D}\right] + \mathfrak{s} = \text{curl } \mathfrak{H} + \frac{1}{c}(\mathfrak{v}\nabla)\mathfrak{D}$$

---

[1] [1] In his letter Mr. Laue gave the correct boundary conditions and pointed out a different derivation of them.

[2] *loc. cit.*

and assume that the current density $\mathfrak{s}$ is finite even in the boundary layer, then the left-hand side of this equation is finite in the boundary layer. The same holds therefore for the right-hand side of the equation.

For an easy interpretation of this result we imagine that the coordinate system is oriented in such a way that a given infinitesimally small portion of the boundary surface, which we now wish to consider, is parallel to the *YZ*-plane. It is then clear that the derivatives of all quantities with respect to y and z remain finite in the above portion of the boundary surface. Hence the totality of those terms on the right-hand side of the above equation that contain differentiations with respect to x must also yield something finite. By simple expansion of the right-hand side and neglecting the terms differentiated with respect to y and z, one arrives at the result that the expressions

$$\frac{\mathfrak{v}_x}{c}\frac{\partial \mathfrak{D}_x}{\partial x},$$

$$\frac{\partial \mathfrak{H}_z}{\partial x} - \frac{\mathfrak{v}_x}{c}\frac{\partial \mathfrak{D}_y}{\partial x},$$

$$\frac{\partial \mathfrak{H}_y}{\partial x} + \frac{\mathfrak{v}_x}{c}\frac{\partial \mathfrak{D}_z}{\partial x},$$

remain finite in the boundary layer. If we also assume that the velocity components of the boundary surface do not experience any jump, then it follows from this that the expressions

$$\mathfrak{D}_x,$$

$$\mathfrak{H}_y + \frac{\mathfrak{v}_x}{c}\mathfrak{D}_z$$

$$\mathfrak{H}_z - \frac{\mathfrak{v}_x}{c}\mathfrak{D}_y$$

have the same value on both sides of the boundary surface (*YZ*-plane). Since $\mathfrak{D}_x$ and the components of $\mathfrak{v}$ are continuous, we can replace the last two expressions by

$$\mathfrak{H}_y - \frac{1}{c}(\mathfrak{v}_z\mathfrak{D}_x - \mathfrak{v}_x\mathfrak{D}_z) ,$$

$$\mathfrak{H}_z - \frac{1}{c}(\mathfrak{v}_x\mathfrak{D}_y - \mathfrak{v}_y\mathfrak{D}_x) .$$

We get rid of the dependence on the special choice of the position of the coordinate axes relative to the boundary surface element considered by writing the result using the notation of vector analysis. If the subscripts n and $\bar{n}$, respectively, denote the components of the pertinent vector in the direction of and perpendicular to the normal of the surface of discontinuity, then it follows that

$$\mathfrak{D}_n ,$$

$$\left\{\mathfrak{H} - \frac{1}{c}[\mathfrak{v}\mathfrak{D}]\right\}_{\bar{n}}$$

must be continuous at the boundary surface.

In the same way one concludes from equation (2a)[1] that the components

$$\mathfrak{B}_n ,$$

$$\left\{\mathfrak{E} + \frac{1}{c}[\mathfrak{v}\mathfrak{B}]\right\}_{\bar{n}}$$

are continuous.

Bern and Würzburg, November 1908. (Received on 6 December 1909)

---

[1] *loc. cit.*

Addendum. If there is a layer of true electricity $(\int \rho \, d\tau)$ of surface density η on the boundary surface under consideration, then $\mathfrak{s}$ becomes infinite. In that case

$$\operatorname{curl} \mathfrak{H} + \frac{1}{c}(\mathfrak{v}\Delta)\mathfrak{D} - \mathfrak{s}$$ [4]

is finite in the boundary layer, where $\mathfrak{s}$ can be replaced by $(\mathfrak{v}/c)\rho$. The above boundary conditions are obtained for this case too, except that the first of those is to be replaced by

$$\mathfrak{D}_{n2} - \mathfrak{D}_{n1} = \eta \ .$$

(Received on 19 January 1909)

Doc 55

COMMENT ON THE PAPER OF D. MIRIMANOFF "ON THE FUNDAMENTAL EQUATIONS..."

by A. Einstein

[*Annalen der Physik* 28 (1909): 885-888]

1. The system of differential equations and transformation equations presented in this paper[1] does not differ from that of Minkowski in any way, [2] or, rather, differs *only* insofar as the vector usually denoted by $\mathfrak{H}$ (magnetic force) was denoted by the author by

$$\mathfrak{Q} = \mathfrak{H} - \frac{1}{c}[\mathfrak{P}\mathfrak{w}] \ .$$

I.e., with the introduction of $\mathfrak{Q}$, as the author himself shows, the [3] differential equation (I) becomes identical with the corresponding equation of

---

[1] D. Mirimanoff, *Ann. d. Phys.* 28 (1909): 192. [1]

Minkowski, while the remaining three differential equations do not contain $\mathfrak{H}$ and already have the form of the corresponding equations of Minkowski. Indeed, the author says himself that his vectors $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{Q}$, and $\mathfrak{B}$ transform like the vectors usually denoted by $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$.

2. Similarly, the relations between the vectors containing material constants (ϵ, μ and σ) do not differ from Minkowski's corresponding relations. I.e., the author starts from the postulate that for a coordinate system instantaneously at rest relative to the system point under consideration, the equations

$$\mathfrak{D} = \epsilon\mathfrak{E}, \qquad \mathfrak{H} = \frac{1}{\mu}\mathfrak{B}, \qquad \mathfrak{J} = \sigma\mathfrak{E}$$

should hold; if one bears in mind that the (author's) vector $\mathfrak{H}$ is identical with the vector $\mathfrak{Q}$ for $\mathfrak{w} = 0$, and that $\mathfrak{Q}$ plays exactly the same role in the author's differential equations and in his transformation equations as $\mathfrak{m}$ does in Minkowski's equations (usually denoted by $\mathfrak{H}$), then one realizes that these equations, too, agree with Minkowski's corresponding equations, except that the notation $\mathfrak{H}$ is replaced by the notation $\mathfrak{Q}$.

3. Thus, it has been shown that Mirimanoff's quantity $\mathfrak{Q}$ plays the same role in all his equations as the quantity usually denoted by $\mathfrak{H}$ and called "magnetic force" or "magnetic field strength." Nevertheless, Mirimanoff's equations would have a different content than those of Minkowski if by definition the quantity $\mathfrak{Q}$ of Mirimanoff would have a different physical meaning than the quantity usually denoted by $\mathfrak{H}$.

In order to reach a conclusion in that matter, we first ask for the meaning of the vectors $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$ in Minkowski's equations

$$\text{(A)} \qquad \left\{ \begin{aligned} &\operatorname{curl} \mathfrak{H} = \frac{1}{c}\frac{\partial \mathfrak{D}}{\partial t} + \mathrm{i}\,, \\ &\operatorname{curl} \mathfrak{E} = -\frac{1}{c}\frac{\partial \mathfrak{B}}{\partial t}\,, \\ &\operatorname{div} \mathfrak{D} = \rho, \\ &\operatorname{div} \mathfrak{B} = 0\,. \end{aligned} \right.$$

One has to admit that these vectors have not yet been expressly defined for the case that the velocity $\mathfrak{w}$ of matter differs from zero; only for the case that $\mathfrak{w}$ vanishes do we possess definitions on which (ideal) measurements of these quantities could be based, and here I have in mind those definitions that are well known from the electrodynamics of bodies at rest. Therefore, if upon using Minkowski's equations we find that in a certain volume element of the body moving with velocity $\mathfrak{w}$ the field vectors at a certain time have certain (vector) values $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$, then we must first transform these field vectors to a reference system that is at rest with respect to the volume element in question. Only the vectors $\mathfrak{E}'$, $\mathfrak{D}'$, $\mathfrak{H}'$, $\mathfrak{B}'$ thus obtained have a definite physical meaning which is known from the electrodynamics of bodies at rest.

Thus, Minkowski's differential equations by themselves do not have any content at the points in which $\mathfrak{w} \neq 0$; however, they do so when taken together with Minkowski's transformation equations and with the stipulation that for the case $\mathfrak{w} = 0$ the definitions of the electrodynamics of bodies at rest must be valid for the field vectors.

We now have to ask: Is Mirimanoff's vector $\mathfrak{Q}$ defined in a different way from the vector we have denoted by $\mathfrak{H}$? This is not the case, for the following reasons:

1. The same differential equations and transformation equations hold for Mirimanoff's field vectors $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{Q}$, $\mathfrak{B}$ as for the vectors $\mathfrak{E}$, $\mathfrak{D}$, $\mathfrak{H}$, $\mathfrak{B}$ of Minkowski's equations (A).

2. Mirimanoff's vector $\mathfrak{Q}$ as well as the vector $\mathfrak{H}$ of (A) are defined only for the case $\mathfrak{w} = 0$. However, in that case, because of Mirimanoff's equation

$$\mathfrak{Q} = \mathfrak{H} - \frac{1}{c}[\mathfrak{P}\mathfrak{w}] ,$$

one has to put $\mathfrak{Q} = \mathfrak{H}$ = field strength; in exactly the same way, in the case $\mathfrak{w} = 0$, the vector $\mathfrak{H}$ of equations (A) is equivalent to the field strength in the sense of the electrodynamics of bodies at rest.

It follows from these two arguments that Mirimanoff's vector $\mathfrak{Q}$ and the vector $\mathfrak{H}$ of (A) are completely equivalent.

4. In order to compare his results regarding Wilson's arrangement with those obtained by Mr. Laub and me, the author should have carried his considerations far enough to arrive at relations between defined quantities, i.e., quantities accessible to observation at least in principle. For this purpose [4]

he would only have had to apply the boundary conditions corresponding to his system of equations. According to what we said above, he would have had to come to the same conclusions as we did, since his theory is identical with Minkowski's.

In conclusion, I would also like to point to the importance of the
[5] recently published paper by Ph. Frank,[1] which, by taking into account the Lorentz contraction, restores the agreement between Lorentz's treatment, based
[6] on the electron theory, and Minkowski's treatment of the electrodynamics of moving bodies.The advantage of the treatment based on the electron theory consists, on the one hand, in providing a graphic interpretation of the field vectors and, on the other hand, in dispensing with the arbitrary assumption that the derivatives of the velocity of matter do not appear in the
[7] differential equations.

[8]
[9] Bern, January 1909. (Received on 22 January 1909)

---

[1] Ph. Frank, *Ann. d. Phys.* 27 (1908): 1059.

Doc. 56

ON THE PRESENT STATUS OF THE RADIATION PROBLEM

by A. Einstein

[*Physikalische Zeitschrift* 10 (1909): 185-193]

This journal has recently published expressions of opinion by Messrs. H. A. Lorentz[1], Jeans[2], and Ritz[3] which offer good insight into the present status of this extremely important problem. In the belief that it would be of benefit if all those who have seriously thought about this matter would communicate their views, even if they have not been able to arrive at a final result, I would like to communicate the following. [4]

1. The simplest form in which we can express the laws of electrodynamics established so far is that presented by the Maxwell-Lorentz partial differential equations. In contrast to Mr. Ritz[3], I regard the forms containing retarded functions as merely auxiliary mathematical forms. The reason I see myself compelled to take this view is first of all that those forms do not subsume the energy principle, while I believe that we should adhere to the strict validity of the energy principle until we shall have found important reasons for renouncing this guiding star. It is certainly true that Maxwell's equations for empty space, taken by themselves, do not say anything, that they only represent an intermediary construct; but, as is well known, exactly the same could be said about Newton's equations of motion, as well as about any theory that needs to be supplemented by other theories in order to yield a picture for a complex of phenomena. What distinguishes the Maxwell-Lorentz differential equations from the forms that contain retarded functions is the circumstance that they yield an expression for the energy and the momentum of the system under consideration for any instant of time, relative to any unaccelerated coordinate system. With a theory that operates with retarded forces it is not possible to describe the instantaneous state of a system at all without using earlier states of the system for this description. [5]

---

[1] H. A. Lorentz, *Phys. Zeit.* 9 (1908): 562-563. [1]

[2] J. H. Jeans, *Phys. Zeit.* 9 (1908): 853-855. [2]

[3] W. Ritz, *Phys. Zeit.* 9 (1908): 903-907. [3]

For example, if a light source A had emitted a light complex toward the screen B, but it has not yet reached the screen B, then, according to theories operating with retarded forces, the light complex is represented by nothing except the processes that have taken place in the emitting body during the preceding emission. Energy and momentum--if one does not want to renounce these quantities altogether--must then be represented as time integrals.

To be sure, Mr. Ritz claims that experience forces us to abandon these differential equations and introduce the retarded potentials. However, his arguments do not seem valid to me.

If one puts with Ritz

$$f_1 = \frac{1}{4\pi} \int \frac{\varphi\left[(x',y',z',t - \frac{r}{c}\right]}{r} dx',dy',dz'$$

and

$$f_2 = \frac{1}{4\pi} \int \frac{\varphi\left[(x',y',z',t + \frac{r}{c}\right]}{r} dx',dy',dz' ,$$

then f_1 as well as f_2 are solutions of the equation

$$\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \Delta f = \varphi(x\ y\ z\ t) ,$$

hence

$$f_3 = a_1 f_1 + a_2 f_2$$

is also a solution if $a_1 + a_2 = 1$. But it is not true that the solution f_3 is a *more general* solution than f_1 and that one specializes the theory by putting $a_1 = 1$, $a_2 = 0$. Putting

$$f(x,y,z,t) = f_1 ,$$

amounts to calculating the electromagnetic effect at the point x,y,z from those motions and configurations of the electric quantities that took place *prior to* the instant t. Putting

$$f(x,y,z,t) = f_2 ,$$

we are determining the above electromagnetic effects from the motions and configurations that take place *after* the instant t.

In the first case the electric field is calculated from the totality of the processes producing it, and in the second case from the totality of the processes absorbing it. If the whole process occurs in a (finite) space bounded on all sides, then it can be represented in the form

$$f = f_1$$

as well as in the form

$$f = f_2 .$$

If we consider a field that is emitted from the finite into the infinite, we can, naturally, use only the form

$$f = f_1 ,$$

precisely because the totality of the absorbing processes *is not taken into consideration*. But here we are dealing with a misleading paradox of the infinite. Both kinds of representation can always be used, regardless of how distant the absorbing bodies are imagined to be. Thus, one cannot conclude that the solution $f = f_1$ is more special than the solution $a_1 f_1 + a_2 f_2$, where $a_1 + a_2 = 1$.

That a body does not "receive energy from infinity unless another body loses a corresponding quantity of energy" cannot be brought up as an argument [6] either, in my opinion. First of all, if we want to stick to experience, we cannot speak of infinity but only of spaces lying outside the space considered. Furthermore, it is no more permissible to infer irreversibility of the electromagnetic *elementary* processes from the nonobservability of such a process than it is permissible to infer irreversibility of the elementary processes of atomic motion from the second law of thermodynamics. [7]

2. Jeans' interpretation can be disputed on the grounds that it might not be permissible to apply the general results of statistical mechanics to cavities filled with radiation. However, the law deduced by Jeans can also be [8] arrived at in the following way[1].

---

[1] Cf. A. Einstein, *Ann. d. Phys.* 17 (1905): 133-136. [9]

According to Maxwell's theory, an ion capable of oscillating about an equilibrium position in the direction of the X-axis will, on the average, emit and absorb equal amounts of energy per unit time only if the following relation holds between the mean oscillation energy $\overline{E_\nu}$ and the energy density of the radiation ρ_ν at the proper frequency ν of the oscillator:

[10]
$$E_\nu = \frac{c^3}{8\pi\nu^2}\rho_\nu , \tag{I}$$

where c denotes the speed of light. If the oscillating ion can also interact with gas molecules (or, generally, with a system that can be described by means of the molecular theory), then we must necessarily have, according to the statistical theory of heat,

$$\overline{E_\nu} = \frac{RT}{N} \tag{II}$$

(R = gas constant, N = number of atoms in one gram-atom, T = absolute temperature), if, on the average, no energy is transferred by the oscillator from the gas to the radiation space[1].

From these two equations we arrive at

$$\rho_\nu = \frac{R}{N}\frac{8\pi}{c^3}\nu^2 T, \tag{III}$$

i.e., exactly the same law that has also been found by Messrs. Jeans and H. A. [12] Lorentz[2].

3. There can be no doubt, in my opinion, that our current theoretical views inevitably lead to the law propounded by Mr. Jeans. However, we can consider it as almost equally well established that formula (III) is not

---

[1] M. Planck, *Ann. d. Phys.* 1 (1900): 99. M. Planck, *Vorlesungen über die* [11] *Theorie der Wärmestrahlung* [Lectures on the theory of thermal radiation], Chapter 3.

[2] It should be explicitly noted that this equation is an inevitable consequence of the statistical theory of heat. The attempt, on p. 178 of the book by [13] Planck just cited, to question the general validity of Equation II, is based, it seems to me, only on a gap in Boltzmann's considerations, which has been [14] filled in the meantime by Gibbs' investigations.

compatible with the facts. Why, after all, do solids emit visible light only above a fixed, rather sharply defined temperature? Why are ultraviolet rays not swarming everywhere if they are indeed constantly being produced at ordinary temperatures? How is it possible to store highly sensitive photographic plates in cassettes for a long time if they constantly produce short-wave rays? For further arguments I refer to §166 of Planck's repeatedly [15] cited work. Thus, we will indeed have to say that experience forces us to [16] reject either equation (I), required by the electromagnetic theory, or equation (II), required by statistical mechanics, or both equations.

4. We must now ask ourselves how Planck's radiation theory relates to the theory which is indicated under 2., and which is based on our currently accepted theoretical foundations. In my opinion the answer to this question is made harder by the fact that Planck's presentation of his own theory suffers from a certain logical imperfection. I will now try to explain this briefly.

a) If one adopts the standpoint that the irreversibility of the processes in nature is only *apparent*, and that the irreversible process consists in a transition to a more probable state, then one must first give a definition of the probability W of a state. The only definition worthy of consideration, in my opinion, would be the following. [17]

Let $A_1, A_2 \ldots A_\ell$ be all the states a closed system at a certain energy content can assume, or, more accurately, all the states that we can distinguish in such a system with the help of certain auxiliary means. According to [18] the classical theory, after a certain time the system will assume one particular state (e.g., A_ℓ) and then remain in this state (thermodynamic equilibrium). However, according to the statistical theory the system will keep assuming, in an irregular sequence, all these states $A_1 \ldots A_\ell$.[1] If the system is observed over a very long time period θ, there will be a certain portion τ_ν of this time such that during τ_ν, and during τ_ν only, the system occupies the state A_ν. [The quantity] $\bar{\tau}_\nu/\theta$ will have a definite limiting [20] value, which we call the probability W of the state A_ν under consideration.

---

[1] That only this last interpretation is tenable follows immediately from the properties of Brownian motion. [19]

Proceeding from this definition, one can show that the entropy S must [21] satisfy the equation

$$S = \frac{R}{N} \lg W + \text{const.},$$

where the constant is the same for all states of the same energy.

b) Neither Mr. Boltzmann nor Mr. Planck gave a definition of W.

They put purely formally W = number of complexions of the state under [22] consideration.

If one now demands that these complexions be equally probable, where the probability of the complexion is defined in the same way that we have defined the probability of the state under (a), one will obtain precisely the definition for the probability of a state given under (a); however, the logically unnecessary element "complexion" has been used in the definition.

Even though the indicated relation between S and W is valid only if the probability of a complexion is defined in the manner indicated or in a manner equivalent to it, neither Mr. Boltzmann nor Mr. Planck has defined the probability of a complexion. But Mr. Boltzmann did clearly realize that the molecular-theoretical picture he had chosen dictated his choice of complexions in a quite definite manner; he discussed this on pages 404 and 405 of his paper "Über die Beziehung..." ["On the relation..."] that appeared in the [23] *Wiener Sitzungsberichte* in 1877.[1] Similarly, Mr. Planck would have had no freedom in the choice of complexions in the resonator theory of radiation. He could have been permitted to postulate the pair of equations

$$S = \frac{R}{N} \lg W$$

and

$$W = \text{number of complexions}$$

only if he had appended the condition that the complexions must be chosen such that in the theoretical model chosen by him they had been found to be equally probable on the basis of statistical considerations. In this way he would [25] have arrived at the formula defended by Jeans. Though every physicist must

---

[1] Cf. L. Boltzman, *Vorlesungen über Gastheorie* (Lectures on the theory of [24] gases), Vol. I, p. 40, lines 9-23.

rejoice that Mr. Planck disregarded these requirements in such a fortunate manner, it should not be forgotten that the Planck radiation formula is incompatible with the theoretical foundation from which Mr. Planck started out.

5. It is simple to see the way in which one could modify the foundations of the Planck theory in order to have the Planck radiation formula truly result from the theoretical foundations. I will not present the pertinent derivations here but will rather just refer to my papers on this subject.[1] The result is as follows: One arrives at the Planck radiation formula if one

1. adheres to equation (I) between resonator energy and radiation pressure, which Planck derived from Maxwell's theory[2];
2. modifies the statistical theory of heat by the following assumption: A structure that is capable of carrying out oscillations with the frequency ν, and which, due to its possession of an electric charge, is capable of converting radiation energy into energy of matter and vice versa, cannot assume oscillation states of any arbitrary energy, but rather only such oscillation states whose energy is a multiple of $h\cdot\nu$. Here h is the constant so designated by Planck, which appears in his radiation equation. [28]

6. Since the modification of the foundations of Planck's theory just described necessarily leads to very profound changes in our physical theories, it is very important to search for the simplest possible, mutually independent interpretations of Planck's radiation formula as well as of the radiation law in general, insofar as the latter may be assumed to be known. Two considerations on this matter, which are distinguished by their simplicity, shall be briefly described below.

Until now, the equation $S = \frac{R}{N} \lg W$ has been applied mainly to calculate the quantity W on the basis of a more or less complete theory, and then to calculate the entropy from W. However, this equation can also be applied conversely, using empirically obtained entropy values S_ν to obtain the

---

[1] A. Einstein, *Ann. d. Phys.* 20 (1906) and *Ann. d. Phys.* 22 (1907), §1. [26]

[2] This amounts to the same as assuming that the electromagnetic theory of radiation at least yields correct time averages. This, however, can hardly be doubted, given the utility of this theory in optics. [27]

statistical probability of the individual states A_ν of an isolated system. A theory yielding values for the probability of a state that differ from those obtained in this way must obviously be rejected.

A consideration of the kind indicated for determining certain statistical properties of heat radiation enclosed in a cavity had already been carried out by me in an earlier paper,[1] in which I first presented the theory of light quanta. However, since at that time I started out from Wien's radiation
[30] formula, which is valid only in the limit (for small values of $\frac{\nu}{T}$), I shall present here a similar consideration which provides a simple interpretation of the content of Planck's radiation formula.

Let V and v be two interconnected spaces bounded by diffusely, completely reflecting walls. Let a heat radiation with the frequency range $d\nu$ be enclosed in these spaces. H shall be the radiation energy existing instantaneously in V, and η the radiation energy existing instantaneously
[31] in v. After some time the proportion $H_0 : \eta_0 = V : v$ will then hold permanently, within some approximation. At an arbitrarily chosen instant of time, η will deviate from η_0 according to a statistical law that is obtained directly from the relation between S and W if one changes over to the differentials,

$$dW = \text{const.}\ e^{\frac{N}{R}\cdot S}\, d\eta\ .$$

If Σ and σ denote the entropy of the radiation in the two respective spaces, and if we set $\eta = \eta_0 + \epsilon$, we have

$$d\eta = d\epsilon$$

and

$$S = \Sigma + \sigma = \Sigma_0 + \sigma_0 + \left\{\frac{d(\Sigma + \sigma)}{d\epsilon}\right\}_0 \epsilon + \frac{1}{2}\left\{\frac{d^2(\Sigma + \sigma)}{d\epsilon^2}\right\}_0 \epsilon^2 \ldots .$$

Because

$$\left\{\frac{d(\Sigma + \sigma)}{d\epsilon}\right\}_0 = 0\ ,$$

---

[29] [1] *Ann. d. Phys.* 17 (1905): 132-148.

if one assumes that V is very large compared with v, the second equation reduces to

$$S = \text{const.} + \frac{1}{2}\left\{\frac{d^2\sigma}{d\epsilon^2}\right\}_0 \epsilon^2 + \ldots .$$ [32]

If we content ourselves with the first nonvanishing term of the series, thus causing an error that is the smaller the greater the value of v compared with the cube of the radiation wavelength, we obtain

$$dW = \text{const.} \cdot e^{-\frac{1}{2}\frac{N}{R}\left[\frac{d^2\sigma}{d\epsilon^2}\right]_0 \epsilon^2} \cdot d\epsilon \,.$$ [33]

From this we obtain for the mean value $\overline{\epsilon^2}$ of the square of the energy fluctuation of the radiation occurring in v

$$\epsilon^2 = \frac{1}{\frac{N}{R}\left\{\frac{d^2\sigma}{d\epsilon^2}\right\}_0} \,.$$ [34]

If the radiation formula is known, we can calculate σ from it.[1] If one considers Planck's radiation formula as an expression of experience, one obtains, after a simple calculation,

$$\epsilon^2 = \frac{R}{Nk}\left\{\nu h \eta_0 + \frac{c^3}{8\pi\nu^2 d\nu} \cdot \frac{\eta_0{}^2}{v}\right\} \,.$$ [36]

We have thus arrived at an easily interpretable expression for the mean value of the fluctuations of the radiation energy present in v. We shall now show that the current theory of radiation is incompatible with this result.

According to the current theory, the fluctuations are due solely to the circumstance that the infinitely many rays traversing the space, which constitute the radiation present in v, interfere with one another and thus provide a momentary energy that is sometimes greater, sometimes smaller than the sum of the energies that the individual rays would provide if they were

---

[1] Cf., e.g., Planck's repeatedly cited book, Equation (230). [35]

not interfering with each other at all. We could thus exactly determine the quantity $\overline{\epsilon^2}$ by a consideration that is mathematically somewhat complicated. We shall content ourselves here with a simple dimensional consideration. The following conditions must be satisfied:

1. The magnitude of the mean fluctuation depends only on λ (wavelength), $d\lambda$, σ, and v, where σ denotes the radiation density related to
[37] the wavelengths $(\sigma d\lambda = \rho d\nu)$.

2. Since the radiation energies of adjacent wavelength ranges and volumes[1] are simply additive, and the corresponding fluctuations are independent of each other, at a given λ and ρ, $\overline{\epsilon^2}$ must be proportional to the quantities $d\lambda$ and v.

3. $\overline{\epsilon^2}$ has the dimension of the square of an energy.

The expression for $\overline{\epsilon^2}$ is thereby completely determined up to a numerical factor (of order of magnitude 1). In this way one arrives at the expression $\sigma^2\lambda^4 v d\lambda$, which upon introduction of the variables used above
[38] reduces to the second term of the formula for $\overline{\epsilon^2}$ just developed. But we would have obtained solely this second term for $\overline{\epsilon^2}$ had we started out with the Jeans formula. One would then also have to put $\frac{R}{Nk}$ equal to a constant of order of magnitude 1, which corresponds to Planck's determination of the
[39] elementary quantum[2]. Thus, the first term of the above expression for $\overline{\epsilon^2}$, which for the visible radiation surrounding us everywhere makes a far greater contribution than the second one, is not compatible with the current theory.

If one would put, with Planck, $\frac{R}{Nk} = 1$, then the first term, if present alone, would yield a fluctuation of the radiation energy equal to that produced if the radiation consisted of point quanta of energy $h\nu$ moving
[40] independently of each other. This can be shown by a simple calculation. One should remember that the contribution of the first term to the average percent fluctuation of energy

$$\left[\sqrt{\frac{\overline{\epsilon^2}}{\overline{\eta_0^2}}}\right]$$

---

[1] Only if these are large enough, of course.

[2] By carrying out the interference consideration indicated above, one would obtain $\frac{R}{Nk} = 1$.

is the greater the smaller the energy η_0, and that the magnitude of this percent fluctuation yielded by the first term is independent of the size of the space v over which the radiation is distributed; I mention this in order to show how fundamentally different the actual statistical properties of radiation are from those to be expected on the basis of our current theory, which is based on linear, homogeneous differential equations. [41]

7. In the foregoing we have calculated the fluctuations of the energy distribution in order to obtain information on the nature of thermal radiation. In what follows we shall briefly show how one can obtain analogous results by calculating the fluctuations of the radiation *pressure*, due to fluctuations of the momentum. [42]

Let a cavity surrounded on all sides by matter of absolute temperature T contain a mirror that can move freely in the direction perpendicular to its normal[1]. If we imagine it to be moving with a certain velocity from the outset, then, due to this motion, more radiation will be reflected at its front than at its back; hence, the radiation pressure acting on the front will be greater than that acting on the back. Thus, due to its motion relative to the cavity radiation, the mirror will be acted upon by a force comparable to friction, which little by little would have to consume the momentum if there did not exist a cause of motion exactly compensating on the average for the momentum lost through the above-mentioned frictional force. To the irregular fluctuations of the energy of a radiation space studied above, there also correspond irregular fluctuations of the momentum, or irregular fluctuations of the pressure forces exerted by the radiation on the mirror, which would have to set the mirror in motion even if it had originally been at rest. The mean speed of the motion of the mirror has then to be determined from the entropy-probability relation, and the law of the above-mentioned frictional forces from the radiation law, which is assumed to be known. From these two results one then calculates the effect of the pressure fluctuations, which in turn makes it possible to draw conclusions concerning the constitution of the radiation or—more precisely—concerning the elementary processes of the reflection of the radiation from the mirror.

---

[1] The motions of the mirror considered here are completely analogous to the so-called Brownian motion of suspended particles.

Let v denote the velocity of the mirror at time t. Owing to the frictional force mentioned above, this velocity decreases by $\frac{Pv\tau}{m}$ in the small time interval τ, where m denotes the mass of the mirror and P the retarding force corresponding to unit velocity of the mirror. Further, we denote by Δ the velocity change of the mirror during τ corresponding to the irregular fluctuations of the radiation pressure. The velocity of the mirror at time $t + \tau$ is

$$v - \frac{P\tau}{m} v + \Delta .$$

For the condition that *on the average* v shall remain unchanged during τ, we obtain

[43] $$\overline{\left[v - \frac{P\tau}{m} v + \Delta\right]} = \overline{v^2}$$

or, if we omit relatively infinitesimal quantities and take into account that the average value of $v\Delta$ obviously vanishes:

[44] $$\Delta^2 = \frac{2P\tau}{m} \overline{v^2} .$$

In this equation $\overline{v^2}$ can be replaced using the equation

[45] $$\frac{\overline{mv^2}}{2} = \frac{1}{2} \frac{RT}{N} ,$$

which can be derived from the entropy-probability equation. Before giving the value of the friction constant P, we specialize the problem under consideration by assuming that the mirror completely reflects the radiation of a certain frequency range (between ν and $\nu + d\nu$) and is completely transparent to radiation of other frequencies. By a calculation omitted here for the
[46] sake of brevity, one obtains from a purely electrodynamic investigation the following equation, which is valid for any arbitrary radiation distribution:

[47] $$P = \frac{3}{2c} \left[\rho - \frac{1}{3} \nu \frac{d\rho}{d\nu}\right] d\nu f ,$$

where ρ again denotes the radiation density at frequency ν, and f the surface area of the mirror. By substituting the values obtained for $\overline{v^2}$ and P, we get

$$\frac{\overline{\Delta^2}}{\tau} = \frac{RT}{N} \cdot \frac{3}{c} \left[\rho - \frac{1}{3} \nu \frac{d\rho}{d\nu}\right] d\nu f .$$

If we transform this expression using Planck's radiation formula, we obtain

$$\frac{\overline{\Delta^2}}{\tau} = \frac{1}{c}\left[h\rho\nu + \frac{c^3}{8\pi} \frac{\rho^2}{\nu^2}\right] d\nu f .$$ [48]

The close connection between this relation and the one derived in the last section for the energy fluctuation $(\overline{\epsilon^2})$ is immediately obvious,[1] and exactly analogous considerations can be applied to it. Again, according to the current theory, the expression would be reduced to the second term (fluctuation due to interference). If the first term alone were present, the fluctuations of the radiation pressure could be completely explained by the assumption that the radiation consists of independently moving, not too extended complexes of energy $h\nu$. In this case, too, the formula says that in [49] accordance with Planck's formula the effects of the two causes of fluctuation mentioned act like fluctuations (errors) arising from mutually independent causes (additivity of the terms of which the square of the fluctuation is composed).

8. In my opinion, the last two considerations conclusively show that the constitution of radiation must be different from what we currently believe. It is true that, as the excellent agreement of theory and experiment in optics has proved, our current theory correctly yields the time averages, which alone can be directly observed, but it necessarily leads to laws on thermal properties of radiation that prove to be incompatible with experience if one maintains the entropy-probability relation. The discrepancy between the phenomena and the theory is the more prominent the larger ν and the smaller ρ. At small ρ the temporal fluctuations of the radiation energy of

---

[1] That relation can be written in the form (assuming $\frac{R}{Nk} = 1$)

$$\overline{\epsilon^2} = \left\{h\rho\nu + \frac{c^3\rho^2}{8\pi\nu^2}\right\} v d\nu$$

a given space or of the force of radiation pressure on a given surface are much larger than expected from our current theory.

We have seen that Planck's radiation law can be understood if one uses the assumption that the oscillation energy of frequency ν can occur only in quanta of magnitude $h\nu$. According to the aforesaid, it is not sufficient to assume that radiation can only be *emitted* and *absorbed* in quanta of this magnitude, i.e., that we are dealing with a property of the emitting or
[50] absorbing matter only; considerations 6 and 7 show that the fluctuations in the spatial distribution of the radiation and in the radiation pressure also occur as if the radiation consisted of quanta of the indicated magnitude. Certainly, it cannot be asserted that the quantum theory follows from Planck's radiation law as a *consequence* and that other interpretations are excluded. However, one can assert indeed that the quantum theory provides the simplest interpretation of the Planck formula.

It should be emphasized that the considerations presented would in the main in no way lose their value if it should turn out that Planck's formula is not valid; it is precisely that part of Planck's formula which has been adequately confirmed by experience (the Wien radiation law valid in the limit
[51] for large $\frac{\nu}{T}$) which leads to the theory of the light quantum.

9. The experimental investigation of the consequences of the theory of light quanta is, in my opinion, one of the most important tasks that the experimental physics of today must solve. The results obtained so far can be divided into three groups.

a) There are clues concerning the energy of those elementary processes that are associated with the absorption or emission of radiation of a certain
[52] frequency (Stokes' rule; velocity of cathode rays produced by light or X-rays; cathode luminescence, etc). To this group also belongs the interesting use Mr. Stark has made of the theory of light quanta to elucidate the peculiar energy distribution in the spectrum of a spectral line emitted by canal rays.[1]

The method of deduction is always as follows: If one elementary process produces another one, then the energy of the latter is not larger than that of the former. On the other hand, the energy of one of the two elementary

---

[53] [1] J. Stark, *Phys. Zeit.* 9 (1908): 767.

processes is known (of magnitude $h\nu$) if the latter consists in the absorption or emission of radiation of a specified frequency.

Especially interesting would be the study of exceptions to Stokes' law. [54] In order to explain these exceptions, one has to assume that a light quantum is emitted only when the emission center in question has absorbed two light quanta. The frequency of such an event, and thus also the intensity of the emitted light having a smaller wavelength than the producing one, will in this case have to be proportional to the square of the intensity of the exciting light at weak irradiation (according to the law of mass action), while according to Stokes' rule a proportionality with the first power of the exciting light intensity is to be expected at weak irradiation. [55]

b) If the absorption[1] of each light quantum brings about an elementary process of a certain kind, then $\frac{E}{h\nu}$ is the number of these elementary processes if the quantity of energy E of radiation of frequency ν is absorbed.

Thus, for example, if the quantity E of radiation of frequency ν is absorbed by a gas being ionized, then it is to be expected that $\frac{E}{Nh\nu}$ gram molecules of the gas will be ionized. This relation only appears to presume the knowledge of N; for if Planck's radiation formula is written in the form

$$\rho = a\nu^3 \frac{1}{e^{\frac{\beta\nu}{T}} - 1},$$

then $\frac{E}{R\beta\nu}$ is the number of gram-molecules ionized.

This relation, which I have already presented in my first paper[2] on this subject, has unfortunately remained unnoticed thus far. [58]

c) The results noted in 5 lead to a modification of the kinetic theory of specific heat[3] and to certain relations between the optical and the thermal behavior of bodies.

---

[1] Of course, the analogous consideration holds also conversely for the production of light by elementary processes (e.g., by collisions of ions). [56]

[2] *Ann. d. Phys.* (4)17 (1905): 132-148, §9. [57]

[3] A. Einstein, *Ann. d. Phys.* (4)22 (1907): 180-190 and 800. [59]

10. It seems difficult to set up a theoretical system that interprets the light quanta in a complete fashion, the way our current molecular mechanics in conjunction with the Maxwell-Lorentz theory is able to interpret the radiation formula propounded by Mr. Jeans. That we are only dealing with a *modification* of our current theory, not with its complete ***abolition***, seems already to be implied by the fact that Jeans' law seems to be valid in the limit (for small $\frac{\nu}{T}$). An indication as to how this modification would probably have to be carried out is given by a dimensional consideration carried
[60] out by Mr. Jeans a few years ago, which is extremely important, in my opinion,
[61] and which--modified in some points--I shall now recount in brief.

Imagine that a closed space contains an ideal gas and radiation and ions, and that owing to their charge, the ions are able to mediate an energy exchange between gas and radiation. In a theory of radiation linked with the consideration of this system the following quantities can be expected to play a role, i.e., to appear in the expression to be obtained for the radiation density ρ:

a) the mean energy η of a molecular structure (up to an unnamed numerical factor equal to $\frac{RT}{N}$),

b) the velocity of light c,

c) the elementary quantum ϵ of electricity,

d) the frequency ν.

From the dimension of ρ, by solely considering the dimensions of the four quantities mentioned above, one can then determine in a simple way what the form of the expression for ρ must be. Substituting the value of $\frac{RT}{N}$ for η, we obtain

$$\rho = \frac{\epsilon^2}{c^4}\,\nu^3\psi(\alpha) ,$$

where

[62]
$$\alpha = \frac{R\epsilon^2}{Nc}\,\frac{\nu}{T} ,$$

where ψ denotes a function that remains undetermined. This equation
[63] contains the Wien displacement law, whose validity can hardly still be in doubt. This has to be understood as a confirmation of the fact that apart from the four quantities introduced above, no other quantities having a dimension play a role in the radiation law.

From this we conclude that, except for dimensionless numerical factors that appear in theoretical developments and of course cannot be determined by dimensional considerations, the coefficients $\frac{\epsilon^2}{c^4}$ and $\frac{R\epsilon^2}{Nc}$ appearing in the [64] equation for ρ must be numerically equal to the coefficients appearing in the Planck (or Wien) radiation formula. Since the above nondeterminable dimensionless numerical factors are hardly likely to essentially change the order of magnitude, we can put, as far as the order of magnitude[1] is concerned

$$\frac{h}{c^3} = \frac{\epsilon^2}{c^4} \quad \text{and} \quad \frac{h}{k} = \frac{R}{N}\frac{\epsilon^2}{c} \text{ ,}$$ [65]

hence

$$h = \frac{\epsilon^2}{c} \quad \text{and} \quad k = \frac{N}{R} \text{ .}$$

It is the second of these equation which has been used by Mr. Planck to determine the elementary quanta of matter or electricity. Concerning the [66] expression for h, it should be noted that

$$h = 6 \cdot 10^{-27}$$

and

$$\frac{\epsilon^2}{c} = 7 \cdot 10^{-30} \text{ .}$$

This is three decimal places off the mark. But this may be due to the fact [67] that the dimensionless factors are not known.

The most important aspect of this derivation is that it relates the light quantum constant h to the elementary quantum ϵ of electricity. We should remember that the elementary quantum ϵ is an outsider in Maxwell-Lorentz electrodynamics[2]. Outside forces must be enlisted in order to construct the electron in the theory; usually, one introduces a rigid framework [69]

---

[1] The Planck formula reads

$$\rho = \frac{8\pi h\nu^3}{c^3}\,\frac{1}{e^{\frac{h\nu}{kT}} - 1} \text{ .}$$

[2] Cf. Levi-Civita. "Sur le mouvement etc." [On the motion, etc.], *Comptes Rendus* (1907). [68]

to prevent the electron's electrical masses from flying apart under the influence of their electric interaction. The relation $h = \frac{\epsilon^2}{c}$ seems to me to indicate that the same modification of the theory that will contain the elementary quantum ϵ as a consequence will also contain the quantum structure of radiation as a consequence. The fundamental equation of optics

$$D(\varphi) = \frac{1}{c^2}\frac{\partial^2\varphi}{\partial^2 t} - \left[\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2}\right] = 0$$

will have to be replaced by an equation in which the universal constant ϵ (probably its square) also appears in a coefficient. The equation sought (or the system of equations sought) must be homogeneous in its dimensions. It must remain unchanged upon application of the Lorentz transformation. It [70] cannot be linear and homogeneous. It must—at least if Jeans' law is really valid in the limit of small $\frac{\nu}{T}$—lead to the form $D(\varphi) = 0$ for large amplitudes in the limit.

I have not yet succeeded in finding a system of equations fulfilling these conditions which would have looked to me suitable for the construction [71] of the elementary electrical quantum and the light quanta. The variety of possibilities does not seem so great, however, for one to have to shrink from this task.

Addendum

From what has been said above under 4. in this paper, the reader could easily get an incorrect impression about the standpoint taken by Mr. Planck with regard to his own theory of thermal radiation. I therefore deem it appropriate to note the following.

In his book, Mr. Planck emphasized in several places that his theory should not yet be viewed as something complete and final. At the end of his introduction, for example, he says verbatim: "I find it important, however, to especially emphasize also at this point the fact, as elaborated in greater detail in the last section of the book, that the theory developed here does not claim by any means to be fully complete, even though, as I believe, it affords a feasible approach by which to consider the processes of energy radiation from the same viewpoint as those of molecular motion." [72]

The pertinent discussions in my paper should not be construed as an objection (in the strict sense of the word) against Planck's theory, but rather as an attempt to formulate and apply the entropy-probability principle more rigorously than has been done till now. A more rigorous formulation of this principle was necessary because without it the subsequent elaborations in the paper, in which the molecular structure of radiation was inferred, would not have been adequately substantiated. So that my conception of the principle would not appear as chosen ad hoc, or arbitrary, I had to show why its current formulation has not completely satisfied me.

Bern, January 1909. (Received on 23 January 1909)

Doc. 57

ON THE PRESENT STATUS OF THE RADIATION PROBLEM

[1] by W. Ritz and A. Einstein

[*Physikalische Zeitschrift* 10 (1909): 323-324]

To clarify the differences of opinion that came to light in our respective publications[1], we note the following.

In the special cases in which an electromagnetic process *remains restricted to a finite space*, the process can be represented in the form

$$f = f_1 = \frac{1}{4\pi}\int \frac{\varphi\left[x',y',z',t-\frac{r}{c}\right]}{r}\,dx'dy'dz'$$

as well as in the form

$$f = f_2 = \frac{1}{4\pi}\int \frac{\varphi\left[x',y',z',t+\frac{r}{c}\right]}{r}\,dx'dy'dz'$$

and in other forms.

While Einstein believes that one could restrict oneself to this case without *substantially* limiting the generality of the consideration, Ritz considers this restriction not to be permissible in *principle*. If one takes this standpoint, then experience compels one to consider the representation by means of retarded potentials as the only one possible, if one is inclined to the view that the fact of irreversibility of radiation processes must already find its expression in the fundamental equations. Ritz considers the restriction to the form of retarded potentials as one of the roots of the second law, while Einstein believes that irreversibility is exclusively due to reasons of [3] probability.

Zurich, April 1909. (Received on 13 April 1909)

---

[2] [1] W. Ritz, *Phys. Zeit.* 9 (1908): 903-907, and A. Einstein, *Phys. Zeit.* 10 (1909): 185-193.

Doc. 58

"Discussion" following lecture version of Henry Siedentopf, "On Ultramicroscopic Images" [1]

[*Physikalische Zeitschrift* 10 (1909): 779-780]

Rubens: I would like to ask Dr. Siedentopf whether he possibly has used this elegant method for quantitative measurements of Brownian molecular motion. With Mr. Perrin's procedure such experiments can be used for the determination of Loschmidt's number. [2] [3]

Lecturer: I would leave this to the physicists working in the laboratory. [4]

Rubens: It strikes me that the method offers the great advantage that the quantities in question can be measured in peace and quiet on a photogram. [5]

Lecturer: Let me ask the gentlemen to take this into their own hands. Professionally, I am so busy that I don't have time for that.

Einstein: The main difficulty is the temperature; it cannot be kept constant. This is also the case with the French measurements. [6]

Seddig: The drawback just mentioned by the previous discussant and consisting in changes in temperature of the preparation during observation due to radiation absorption, which affect the result in an uncontrollable fashion, was also experienced by me in similar investigations, and for this reason I did not use continuous illumination in my attempts to follow the Brownian molecular motion quantitatively: instead, I always passed two very short flashes of light through the preparation to mark the instantaneous position of the particles photographically. In fact, a method somewhat similar to the one just described was used by The Svedberg 2 or 3 years ago; he let the colloidal solution flow slowly through the test cuvette, and then he obtained similar deviations from rectilinear motion, which he of course did not photograph but rather observed through an eyepiece micrometer. [7] [8] [9]

Lecturer: The Svedberg's method has the flaw that one cannot control the interfering currents, which even get magnified. Also, he incorrectly assumed the motion to be sinusoidal. [10]

Doc. 59

"Discussion" following lecture version of Arthur Szarvassi, "The Theory of Electromagnetic Phenomena in Moving Bodies and the Energy Principle"

[*Physikalische Zeitschrift* 10 (1909): 813]

[1] *Einstein*: I think that a body subjected to some forces, that this body, when observed from a relatively moving coordinate system, because of its being [2] subjected to forces, represents an energy. If this assumption is not made, there will arise a violation of the energy principle. Don't you think that this may be the basis of the example you considered? Is it clear to you what I mean?

Lecturer: Not completely.

Einstein: It can be shown that a moving body subjected to forces whose resultant does not vanish, in certain cases does not get accelerated thereby. One must therefore assume in the theory of relativity that the moving (rigid) body subjected to forces possesses a certain energy content; otherwise, one gets a violation of the energy principle.

Lecturer: That would mean that, besides the so-called kinetic energy in the currently accepted sense, and besides the so-called usual potential electric energy of the system, there exists an additional part of energy. Because I made no assumption about the quantity ψ, the above part would have to be contained in this function. After all, I did not say anything about the form [3] of the function ψ. The energy equation expresses the energy principle quite generally. It is very likely that this energy quantity is contained in ψ.

Einstein: I cannot comment on that, because I did not enter sufficiently into the spirit of this consideration.

Doc. 60
ON THE DEVELOPMENT OF OUR VIEWS CONCERNING THE NATURE AND CONSTITUTION OF RADIATION

by A. Einstein

[*Deutsche Physikalische Gesellschaft, Verhandlungen* 7 (1909): 482-500.
Also in *Physikalische Zeitschrift* 10 (1909): 817-826]

(Presented at the session of the Division of Physics of the 81st Meeting of German Scientists and Physicians in Salzburg on September 21, 1909.) [1]
(Cf. above p. 417)

Once it had been recognized that light exhibits the phenomena of interference and diffraction, it seemed hardly doubtful any longer that light is to be conceived as a wave motion. Since light can also propagate through vacuum, one had to imagine that vacuum, too, contains some special kind of matter that mediates the propagation of light waves. For the interpretation of the laws of the propagation of light in ponderable bodies, it was necessary to assume that this matter, which was called luminiferous ether, is present in them too, and that in the interior of ponderable bodies as well, it is essentially the luminiferous ether that mediates the propagation of light. The existence of this luminiferous ether seemed beyond doubt. The first volume of the excellent textbook by Chwolson, which was published in 1902, contains in the Introduction the following sentence about the ether: "The probability of the hypothesis on the existence of this agent borders extraordinarily closely on certainty." [2]

However, today we must regard the ether hypothesis as an obsolete standpoint. It is even undeniable that there is an extensive group of facts concerning radiation that shows that light possesses certain fundamental properties that can be understood far more readily from the standpoint of Newton's emission theory of light than from the standpoint of the wave theory. [3] It is therefore my opinion that the next stage in the development of theoretical physics will bring us a theory of light that can be understood as a kind of fusion of the wave and emission theories of light. To give reasons for this opinion and to show that a profound change in our views on the nature and constitution of light is imperative is the purpose of the following remarks.

The greatest progress theoretical optics has made since the introduction of the wave theory consists doubtless in Maxwell's brilliant discovery that light might be viewed as an electromagnetic process. Instead of mechanical quantities, i.e., deformation and velocity of parts of the ether, this theory introduces into consideration the electromagnetic states of ether and matter and thereby reduces optical problems to electromagnetic ones. The more electromagnetic theory advanced, the more the question of whether electromagnetic processes can be reduced to mechanical ones retreated into the background; one became used to considering the concepts of electric and magnetic field strength, electric space density, etc., as elementary concepts [4] that are not in need of mechanistic interpretation.

The introduction of the electromagnetic theory brought about a simplification of the bases of theoretical optics and a reduction in the number of arbitrary hypotheses. The old question about the direction of oscillation of [5] polarized light became moot. The difficulties with boundary conditions at the boundary of two media were resolved by the foundation of the theory. There was no longer a need for an arbitrary hypothesis in order to exclude* longitudinal light waves. The pressure of light, which has only recently been [6] established experimentally, and which plays such an important role in the theory of radiation, proved to be a consequence of the theory. I will not attempt here an exhaustive enumeration of the well-known achievements but will rather consider a cardinal aspect in which the electromagnetic theory agrees with or, more accurately, seems to agree with the kinetic theory.

According to both of these theories, light waves appear to be essentially an aggregate of states of a hypothetical medium, the ether, which is present everywhere even in the absence of radiation. It had therefore to be assumed that the movements of this medium must influence the optical and electromagnetic phenomena. The search for the laws governing this influence has caused a transformation in the fundamental views about the nature of radiation, the course of which we want to consider briefly.

The basic question that arose in this regard was the following: Does the luminiferous ether take part in the motion of matter or does it move inside the moving matter in a different way; or, finally, could it be possible

---

*Translator's note: The German text erroneously says "anzuschliessen" (to connect) instead of "auszuschliessen" (to exclude).

that it does not participate in the motion of matter at all but stays at rest instead? To solve this problem, Fizeau performed an important interference experiment based on the following consideration. Let light propagate in a [7] body with velocity V if the body is at rest. If this body, when in motion, completely carries along its ether, the light will, relative to the body, propagate in the same way as if the body were at rest. Hence the propagation velocity relative to the body will in this case also be V. However, in absolute terms, i.e., relative to an observer not moving along with the body, the propagation velocity of a light ray will equal the geometric sum of V and the velocity of motion v of the body. If the velocities of propagation and of motion have the same direction and the same sense, then V_{abs} simply equals the sum of the two velocities, i.e.,

$$V_{\text{abs}} = V + v \ .$$

To test whether this consequence of the hypothesis of the completely co-moving luminiferous ether is correct, Fizeau made each of two coherent [8] monochromatic beams of light pass axially through one of two water-filled tubes and then interfere with each other. When he then let both the water and the light move axially through the tubes, in the direction of the light in one tube and in the opposite direction in the other tube, he obtained a shift in the interference fringes from which he could draw a conclusion about the effect of the velocity of the body on the absolute velocity.

It turned out, as we know, that the velocity of the body does show an influence in the sense expected, but that this influence is smaller than the hypothesis of complete drag would require. We have

$$V_{\text{abs}} = V + \alpha v \ ,$$

where α is always smaller than 1. Neglecting dispersion, we get

$$\alpha = 1 - \frac{1}{n^2} \ .$$

This experiment showed that the ether is not fully carried along by matter, i.e., that in general a relative motion of the ether with respect to matter is taking place. But the earth is a body that in the course of the year has velocities of varying directions with respect to the solar system, and it had to be assumed that the ether in our laboratories does not completely participate in this motion of the earth, just as it did not seem to participate completely in the motion of water in Fizeau's experiment. Thus, we had to conclude that there exists a relative motion of the ether with respect to our apparatuses that changes with the time of day and with the seasons, and one had to expect that this relative motion would cause an apparent anisotropy of space in optical experiments, i.e., that optical [10] phenomena depend on the orientation of the apparatuses. All kinds of experiments have been carried out to detect such an anisotropy, but the expected dependence on the orientation of the apparatuses could not be established.

This contradiction was resolved for the most part by the trailblazing [11] work of H. A. Lorentz in 1895. Lorentz showed that by assuming an ether that is at rest and does not take part in the motion of matter, one can arrive at a theory that accounts for almost all phenomena without having to postulate other hypotheses. In particular, explanations were obtained for the results of the experiment of Fizeau outlined above and for the negative result of the above-mentioned attempts to demonstrate the motion of the earth relative to the ether. There was only one single test that seemed incompatible with [12] Lorentz's theory, namely the interference experiment of Michelson and Morley.

Lorentz had shown that according to his theory, apart from terms containing as a factor the second or higher power of the ratio $\frac{\text{velocity of the body}}{\text{velocity of light}}$, the common translatory motion of the apparatuses has no effect on the ray paths in optical experiments. But one already knew the interference experiment of Michelson and Morley, which had shown that in one special case the terms containing the second power of the ratio $\frac{\text{velocity of the body}}{\text{velocity of light}}$ are not observed even though this had been expected from the standpoint of the theory of an ether at rest. In order to have this [13] experiment encompassed by the theory, Lorentz and FitzGerald introduced, as we know, the assumption that all bodies, hence also those connecting the components of the experimental arrangement of Michelson and Morley, change their shape in a certain way when they are set in motion relative to the ether.

This state of affairs was extremely unsatisfactory, however. The only theory that was workable and transparent as to its foundations was Lorentz's theory. This theory rested on the assumption of a completely immobile ether. The earth had to be viewed as in motion relative to this ether. However, all attempts to prove this relative motion had ended in failure, so that one was forced to introduce a quite peculiar hypothesis in order to understand why this relative motion was not observable.

Michelson's experiment suggested the assumption that, relative to a coordinate system moving along with the earth, and, more generally, relative to any system in nonaccelerated motion, all phenomena proceed according to exactly identical laws. Henceforth, we will call this assumption in brief "the principle of relativity." Before we consider the question of whether it [14] is possible to adhere to the principle of relativity, we shall briefly discuss what will become of the ether hypothesis if we adhere to this principle.

Taking as a basis the ether hypothesis, the experiment led to the supposition that the ether is immobile. In that case the principle of relativity states that all laws of nature referred to a coordinate system K' that is in uniform motion relative to the ether must be identical with the corresponding laws referred to a coordinate system K that is at rest relative to it. But if this is so, then we have just as much reason to imagine the ether at rest relative to K' as at rest relative to K. Hence it is totally unnatural to single out one of the two coordinate systems K, K' by introducing an ether that is at rest relative to it. From this it follows that one can obtain a satisfactory theory only if one drops the ether hypothesis. In that case the electromagnetic fields that constitute the light will no longer appear to be states of a hypothetical medium, but rather independent entities emitted by the sources of light, exactly as in the Newtonian emission theory of light. Exactly as according to the latter [15] theory, a space not permeated by radiation and free of ponderable matter appears to be really empty.

Upon superficial consideration, it seems impossible to reconcile the gist of Lorentz's theory with the principle of relativity. For according to Lorentz's theory, if a ray of light is propagated in vacuum, this always occurs with the fixed velocity c with respect to a coordinate system K at rest in the ether, independent of the state of motion of the emitting body.

We will call this proposition the principle of the constancy of the velocity of light. According to the addition theorem of velocities, this same ray of light will not also propagate with the velocity c with respect to a coordinate system K' that is in translatory motion relative to the ether. The laws of propagation of light thus seem to be different relative to the two coordinate systems, and from this it seems to follow that the principle of relativity is incompatible with the laws of propagation of light.

However, the addition theorem of velocities is based on the arbitrary assumptions that statements concerning time as well as those concerning the shape of moving bodies have a meaning that is independent of the state of motion of the coordinate system used. One can see, however, that a definition of time and of the shape of moving bodies necessitates the introduction of clocks that are at rest with respect to the coordinate system used. The above concepts must therefore be defined separately for each coordinate system, and it is not a foregone conclusion that for two coordinate systems K and K' moving relative to each other these definitions yield the same time values t and t' for the individual events; it is equally impossible to say a priori that every statement about the shape of the bodies valid with respect to the coordinate system K shall also be valid with respect to the coordinate system K' that is moving relative to K.

From this it follows that the currently used transformation equations for the transition from one coordinate system to another one moving uniformly relative to it are based on arbitrary assumptions. If these are dropped, then the foundations of the Lorentz theory, or, more generally, the principle of the constancy of the velocity of light, turn out to be reconcilable with the principle of relativity. One thus arrives at new equations for the transformation of coordinates, which are uniquely determined by the two principles, and which, given the appropriate choice of the origins of coordinates and times, are characterized by the fact that through them the equation

$$x^2 + y^2 + z^2 - c^2t^2 = x'^2 + y'^2 + z'^2 - c^2t'^2$$

becomes an identity. Here c denotes the velocity of light in vacuum; x, y, z, t are space-time coordinates with reference to K, and x', y', z', t' with reference to K'.

This path leads to the so-called theory of relativity, of whose consequences I would here like to mention only one, because it brings about a certain modification of the basic ideas of physics. It turns out that the inertial mass of a body decreases by L/c^2 when the body emits the radiation energy L. One can arrive at this in the following way. [16]

We consider a motionless, freely floating body that emits in two opposite directions the same amount of energy in the form of radiation. The body remains at rest. If E_0 denotes the energy of the body before the emission, E_1 its energy after the emission, and L the amount of radiation emitted, then we have, according to the energy principle

$$E_0 = E_1 + L \; .$$

We now observe the body, as well as the radiation it emits, from a coordinate system relative to which the body moves with velocity v. The theory of relativity then provides the means for calculating the energy of the emitted radiation with respect to the new coordinate system. The value obtained for it is

$$L' = L \cdot \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \; .$$

Since the principle of conservation of energy must hold for the new coordinate system as well, one obtains, using an analogous notation,

$$E_0' = E_1' + L \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \; .$$

By subtraction, and omitting terms of fourth or higher order in v/c, we get

$$(E_0' - E_0) = (E_1' - E_1) + \frac{1}{2} \frac{L}{c^2} v^2 \; .$$

However, $E_0' - E_0$ is nothing other than the kinetic energy of the body before the emission of light, and $E_1' - E_1$ is nothing other than its kinetic energy after the emission of light. If M_0 denotes the mass of the body before the emission, and M_1 its mass after the emission, then one can put, neglecting terms of higher than the second power,

$$\frac{1}{2} M_0 v^2 = \frac{1}{2} M_1 v^2 + \frac{1}{2} \frac{L}{c^2} v^2$$

or

$$M_0 = M_1 + \frac{L}{c^2} .$$

Thus the inertial mass of a body decreases upon emission of light. The energy emitted must be reckoned as part of the body's mass. From this it can be concluded further that each absorption or release of energy brings about, respectively, an increase or decrease of the mass of the body involved. Energy and mass appear as equivalent quantities the same way that heat and mechanical energy do.

The theory of relativity has thus changed our views on the nature of light insofar as it does not conceive of light as a sequence of states of a hypothetical medium, but rather as something having an independent existence just like matter. Furthermore, this theory shares with the corpuscular theory of light the characteristic feature of a transfer of inertial mass from the emitting to the absorbing body. Regarding our conception of the structure of light, in particular of the distribution of energy in the irradiated space, the theory of relativity did not change anything. It is nevertheless my opinion that with respect to this aspect of the problem we are at the threshold of not yet fully foreseeable, but nevertheless highly significant, developments. What I shall presently say is for the most part my private opinion or, rather, the result of considerations that have not yet been sufficiently checked by others. If I nevertheless present these considerations, this should not be attributed to excessive confidence in my own views but rather to the hope that I may induce one or another among you to concern himself with the problems in question.

Even without getting deeper into any theoretical consideration, one can notice that our theory of light cannot explain certain fundamental properties
[17] of light phenomena. Why does it depend only on the color, but not on the intensity, of light whether a given photochemical reaction will or will not occur? Why are short-wave rays in general chemically more effective than long-wave ones? Why is the velocity of photoelectrically produced cathode

rays independent of the intensity of light? Why does one need high temperatures, i.e., high molecular energies, if the radiation emitted by bodies is to contain short-wave components?

The wave theory in its current form does not answer any of these questions. In particular, it is absolutely incomprehensible why the cathode rays produced photoelectrically or by X-rays attain such a substantial velocity independent of radiation intensity. The appearance of such large [18] amounts of energy on a molecular structure under the influence of a source in which the energy is so sparsely distributed, as we must assume according to the wave theory for light and X-rays, prompted capable physicists to resort to a rather farfetched hypothesis. They assumed that light plays merely a triggering role in the process, and that the molecular energies that come to the fore are of radioactive nature. As this hypothesis has by now been more [19] or less abandoned, I will not present any arguments against it.

The basic property of the wave theory that gives rise to these difficulties seems to me to lie in the following. While in the kinetic theory of matter there exists an inverse process for every process in which only a few elementary particles take part, e.g., for every molecular collision, according to the wave theory this is not the case for elementary radiation processes. According to the prevailing theory, an oscillating ion produces an outwardly propagated spherical wave. The opposite process does not exist as *an elementary process*. It is true that the inwardly propagated spherical wave is mathematically possible; however, its approximate realization requires an enormous amount of emitting elementary structures. Thus, the elementary process of light radiation as such does not possess the character of reversibility. Here, I believe, our wave theory is off the mark. Concerning this point the Newtonian emission theory of light seems to contain more truth than does the wave theory, since according to the former the energy imparted at emission to a particle of light is not scattered throughout the infinite space but remains available for an elementary process of absorption. Keep in mind the laws of production of secondary cathode rays by X-rays.

If primary cathode rays impinge upon a metal plate P_1, they produce X-rays. If these impinge upon a second metal plate P_2, cathode rays will be produced once again, their velocity being of the same order of magnitude as that of the primary cathode rays. As far as we know today, the velocity of [20]

the secondary cathode rays depends neither on the distance between the plates P_1 and P_2 upon which they impinge nor on the intensity of the primary cathode rays, but exclusively on the velocity of the primary cathode rays. Let us for once assume that this is strictly valid. What will happen if we let the intensity of the primary cathode rays, or the size of the plate P_1 upon which they impinge, decrease to such a degree that the impinging of an electron of the primary cathode rays can be conceived as an isolated process? If the above is really true, then, because the velocity of the secondary rays is independent of the intensity of the primary rays, we will have to assume that on P_2 (as a result of the impinging of the above electron on P_1) either nothing is being produced or that a secondary emission of an electron occurs on it with a velocity of the same order of magnitude as of the electron impinging on P_1. In other words, the elementary radiation process seems to proceed such that it does not, as the wave theory would require, distribute and scatter the energy of the primary electron in a spherical wave propagating in all directions. Rather, it seems that at least a large part of this energy is available at some location of P_2 or somewhere else. *The elementary pro-*
[21] *cess of radiation seems to be* ***directed***. Furthermore, one gets the impression that the process of X-ray production in P_1 and the process of secondary cathode ray production in P_2 are essentially inverse processes.

The constitution of radiation thus seems to be different from that following from our wave theory. Important clues to that effect have been provided by the theory of temperature radiation, first and foremost by the theory on which Mr. Planck has based his radiation formula. Since I cannot assume that this theory is universally known, I will briefly describe its
[22] essential points.

The interior of a cavity of temperature T contains radiation whose composition is independent of the nature of the body. The amount of radiation in the cavity , whose frequency lies between ν and $\nu + d\nu$, is $\rho d\nu$ per unit volume. The problem consists of determining ρ as a function of ν and T. If an electric resonator of proper frequency ν_0 and slight attenuation is present in the cavity, the electromagnetic theory of radiation enables us to calculate the time average of the energy ($\bar{E}$) of the resonator as a function of $\rho(\nu_0)$. The problem is thereby reduced to one of determining $\bar{E}$ as a function of the temperature. The latter problem can in turn be reduced to the following. Let the cavity contain a very large number (N) of resonators of frequency ν_0. How does the entropy of this system of resonators depend on the latter's energy?

To solve this problem, Mr. Planck uses the general relation between entropy and the probability of the state as inferred by Boltzmann from his investigations on the theory of gases. We have, in general,

$$\text{entropy} = k \cdot \log W,$$

where k denotes a universal constant and W the ***probability*** of the state under consideration. This probability is measured by the "number of complexions," a number that indicates in how many different ways the state in question can be realized. In the case of the above problem, the state of the resonator system is defined by its total energy, so that the problem to be solved reads: In how many different ways can the given total energy be distributed among N resonators? In order to determine this, Mr. Planck divides the total energy into equal small parts of a certain magnitude ϵ. A complexion is determined by stating how many such ϵ's belong to each resonator. The number of such complexions, which yield the total energy, is determined and set equal to W.

From the Wien displacement law, which can be derived thermodynamically, Mr. Planck then concludes further that one has to set $\epsilon = h\nu$, where h denotes a number that is independent of ν. This way he arrives at his radiation formula

$$\rho = \frac{8\pi h\nu^3}{c^3} \cdot \frac{1}{e^{\frac{h\nu}{kT}} - 1},$$

which fully agrees with experience thus far.

It might seem that according to this derivation the Planck radiation formula has to be viewed as a consequence of the current electromagnetic theory of radiation. However, this is not the case, especially for the following reason. The number of complexions just discussed could be viewed as an expression of the multiplicity of probabilities of distribution of the total energy among N resonators only in case every imaginable distribution of energy would appear, at least to some approximation, among the complexions [23] used for calculating W. This requires that for all ν's to which corresponds a perceptible energy density ρ, the energy quantum ϵ be small compared with the mean energy of the resonators $\bar{E}$. However, one finds by simple calculation that for the wavelength of $0.5\ \mu$ and absolute temperature $T = 1700$, $\epsilon/\bar{E}$ is not only not small compared to 1, but is in fact very large compared to it. Its value is about 6.5×10^7. Thus, in the numerical example presented, we must proceed with the counting of the complexions as though the energy of the resonators could only assume the value zero, the 6.5×10^7-fold of its mean energy value, or a multiple of this. It is clear that if we proceed in this manner, we use for the calculation of the entropy only a vanishingly small part of those energy distributions that we must consider as possible according to the foundations of the theory. Thus, according to the foundations of the theory, the number of these complexions is not an expression for the probability of the state in Boltzmann's sense. In my opinion, to accept Planck's theory means plainly to reject the foundations of our radiation theory.

I have already tried to show that our current foundations of the radiation theory must be abandoned. At any rate, there can be no question of rejecting Planck's theory on the grounds that it does not fit these foundations. This theory led to a determination of the elementary quanta that has been splendidly confirmed by the most recent measurements of these quantities based on the counting of α-particles. For the elementary quantum of electricity Rutherford and Geiger obtained on the average a value of $4.65 \cdot 10^{-10}$
[25] and Regener $4.79 \cdot 10^{-10}$, while Mr. Planck with the aid of his radiation theory obtained from the constants of the radiation formula the intermediate
[26] value $4.69 \cdot 10^{-10}$.

Planck's theory leads to the following conjecture. If it is really true that a radiation resonator can only assume energy values that are a multiple of $h\nu$, then it is logical to assume that emission and absorption of radiation can take place only in quanta of this energy value. On the basis of this
[27] hypothesis, the hypothesis of light quanta, one can answer the questions raised above regarding the absorption and emission of radiation. As far as we know, the quantitative consequences of this hypothesis of light quanta are also being confirmed. The following question arises then. Isn't it conceivable that Planck's formula is correct, but that nevertheless a derivation of

it can be given that is not based on an assumption as horrendous-looking as Planck's theory? Would it not be possible to replace the hypothesis of light quanta by another assumption that would also fit the known phenomena? If it is necessary to modify the elements of the theory, would it not be possible to retain at least the equations for the propagation of radiation and conceive only the elementary processes of emission and absorption differently than they have been until now?

To clarify these matters, we will try to proceed in the opposite direction than that taken by Mr. Planck in his radiation theory. We consider Planck's radiation formula as correct and ask ourselves whether some conclusion about the constitution of radiation can be inferred from it. Of two considerations I have carried out in this sense, I will here outline for you only one which, because of its clarity, seems to me especially persuasive. [28]

Let a cavity contain an ideal gas as well as a plate made of a solid substance that can move freely only perpendicular to its plane. Because of the irregularity of the collisions between the gas molecules and the plate, the latter will be set in motion such that its average kinetic energy equals one-third of the average kinetic energy of a monoatomic gas molecule. This is a conclusion drawn from statistical mechanics. We now assume that besides the gas, which we may conceive as consisting of few molecules, there is also radiation present in the cavity; let this radiation be the so-called temperature radiation having the same temperature as the gas. This will be the case if the walls of the cavity have the definite temperature T, are impermeable to radiation, and are not everywhere completely reflecting toward the cavity. Further, we shall temporarily assume that our plate is completely reflecting on both sides. In this state of affairs, not only the gas but also the radiation will be acting upon the plate. The radiation will exert pressure on both sides of the plate. The forces of pressure exerted on the two sides are equal if the plate is at rest. However, if it is in motion, more radiation will be reflected on the surface that is ahead during the motion (front surface) than on the back surface. The backward-acting force of pressure exerted on the front surface is thus larger than the force of pressure acting on the back. Hence, as the resultant of the two forces, there remains a force that counteracts the motion of the plate and that increases with the velocity of the plate. We will call this resultant "radiation friction" in brief.

Assuming for an instant that the above took full account of the mechanical effect of radiation on the plate, we arrive at the following interpretation. Through collision with gas molecules, impulses of random direction are imparted to the plate at random intervals. The velocity of the plate always decreases between two such collisions because of radiation friction, and kinetic energy of the plate is converted into radiation energy. Consequently the energy of the gas molecules would continually be converted into radiation energy by means of the plate until all the energy had turned into radiation energy. Hence no equilibrium could exist between gas and radiation.

This consideration is faulty because one cannot consider the forces of pressure exerted on the plate by radiation as constant in time and free of random fluctuations,just like the forces of pressure exerted on the plate by the gas. For thermal equilibrium to be possible, the fluctuations of the radiation pressure must be such that on the average they compensate for the velocity losses of the plate caused by radiation friction, where the average kinetic energy of the plate equals one-third of the average kinetic energy of a monoatomic gas molecule. If the law of radiation is known, one can calculate the radiation friction, and from this one can calculate the average value of the momenta imparted to the plate due to fluctuations of the radiation pressure so that statistical equilibrium can exist.

This consideration becomes even more interesting if the plate is chosen such that only radiation of the frequency interval $d\nu$ is completely reflected, while radiation of other frequencies passes through without absorption; then one obtains the fluctuations of the radiation pressure for the radiation in the frequency interval $d\nu$. I shall now present the result of the calculation for this case: If Δ denotes the momentum transferred to the plate during time τ as a result of the irregular fluctuations of the radiation pressure, one obtains for the mean square of Δ the expression

$$\overline{\Delta^2} = \frac{1}{c}\left[h\rho\nu + \frac{c^3}{8\pi}\frac{\rho^2}{\nu^2}\right]d\nu f\tau \ .$$

This expression, first of all, is conspicuous by its simplicity; there is not likely to exist another radiation formula agreeing with experience within the range of experimental error that presents such a simple expression for the statistical properties of the radiation pressure as does that of Planck.

As far as interpretation is concerned, the first thing to note is that the expression for the mean square of fluctuation is a sum of two terms. [29] Thus, it appears that there exist two different, independent factors causing a fluctuation of the radiation pressure. From the fact that $\overline{\Delta^2}$ is proportional to f, we conclude that pressure fluctuations for adjacent parts of the plate, whose linear dimensions are large compared with the wavelengths of the reflection frequency, are mutually independent events.

The wave theory provides an explanation only for the second term of the expression found for $\overline{\Delta^2}$. According to the wave theory, beams of not very [30] different directions, not very different frequencies, and not very different states of polarization must interfere with each other, and to the totality of these interferences, which occur in the most random fashion, there must correspond a fluctuation of the radiation pressure. That the expression for this fluctuation must have the form of the second term of our formula can be seen by a simple dimensional analysis. One can see that the wave structure of radiation indeed causes the fluctuations of radiation pressure to be expected from it.

But how to explain the first term of the formula? This term is by no means to be neglected; on the contrary, it alone is relevant, so to speak, in the domain of validity of the so-called Wien radiation law. For $\lambda = 0.5\ \mu$ and $T = 1700$, for example, this term is about $6.5 \cdot 10^7$ times larger than the second one. If radiation consisted of very small-sized complexes of energy $h\nu$, moving through space independently of each other and reflected independently of each other—a conception that represents the very roughest visualization of the hypothesis of light quanta—then the momenta acting on our plate due to fluctuations of the radiation pressure would be of the kind represented by the first term alone.

Thus, in my opinion, the following must be concluded from the above formula, which is, in turn, a consequence of Planck's radiation formula. *In addition to the nonuniformities in the spatial distribution of the momentum of the radiation which arise from the wave theory, there also exist other nonuniformities in the spatial distribution of the momentum, which at low energy density of the radiation have a far greater influence then the first-mentioned* *nonuniformities*. I shall add that another consideration concerning the spatial distribution of the energy yields results that agree quite well with the above consideration concerning the spatial distribution of the momentum.

As far as I know, it has not yet been possible to formulate a mathematical theory of radiation that would do justice both to the undulatory structure and to the structure inferred from the first term of the above formula (quantum structure). The difficulty lies mainly in the fact that the fluctuation properties of radiation, as expressed by the above formula, offer few formal clues on which to build a theory. Imagine that the diffraction and interference phenomena are not yet known, but that one knows that the average magnitude of the irregular fluctuations of the radiation pressure is determined by the second term of the above formula, where ν is a parameter of unknown significance that determines the color. Who would have sufficient imagination to construct the wave theory of light on such a basis?

Still, for the time being the most natural interpretation seems to me to be that the occurence of electromagnetic fields of light is associated with singular points just like the occurence of electrostatic fields according to
[31] the electron theory. It is not out of the question that in such a theory the entire energy of the electromagnetic field might be viewed as localized in these singularities, exactly like in the old theory of action at a distance.
[32] I more or less imagine each such singular point as being surrounded by a field of force which has essentially the character of a plane wave and whose amplitude decreases with the distance from the singular point. If many such singularities are present at separations that are small compared with the dimensions of the field of force of a singular point, then such fields of force will superpose, and their totality will yield an undulatory field of force that may differ only slightly from an undulatory field as defined by the current electromagnetic theory of light. I am sure it need not be particularly emphasized that no importance should be attached to such a picture as long as it has not led to an exact theory. All I wanted is briefly to indicate with its help that the two structural properties (the undulatory structure and the quantum structure) simultaneously displayed by radiation according to the Planck formula should not be considered as mutually incompatible.

Doc. 61

"Discussion" following lecture version of "On the Development of our Views Concerning the Nature and Constitution of Radiation" [Doc. 60]

[*Physikalische Zeitschrift* 10 (1909): 825-826]

Planck: While I am taking the liberty to say a few words of comment on the lecture, I will begin by joining in the thanks of the entire audience, which listened with the greatest of interest to Mr. Einstein's presentation and was stimulated to further reflection even where perhaps opposition may have emerged. Naturally, I will restrict myself to the things in which my opinion differs from that of the lecturer. After all, most of what the lecturer has been saying will not meet with any disagreement. I, too, emphasize the necessity of introducing certain quanta. We cannot progress with the radiation theory unless we divide, in a certain sense, the energy into quanta, which are to be conceived as atoms of action. The question is now where to look for these quanta. According to the latest considerations of Mr. Einstein, it would be necessary to conceive the free radiation in vacuum, and thus the light waves themselves, as atomistically constituted, and hence to give up Maxwell's equations. This seems to me a step which in my opinion is not yet necessary. I will not go into details, but will rather note the [1] following. In the latest consideration by Mr. Einstein he inferred the fluctuations of free radiation in pure vacuum from the motion of matter. This inference seems to me absolutely irreproachable only in the case that the interactions between the radiation in vacuum and the motion of matter are completely known; if this is not the case, then the bridge necessary to cross from the motion of the mirror to the intensity of the incident light is missing. However, it seems to me that we know very little about this interaction between the free electrical energy in vacuum and the motion of the atoms of matter. This interaction is essentially based on the emission and absorption of light. Essentially this is also the case for radiation pressure, at least according to the generally accepted theory of dispersion, which also reduces reflection to absorption and emission. However, it is just emission and absorption which are the obscure points about which we know very little. We may know a little about absorption, but what about emission? We imagine that it is produced by the acceleration of electrons. But this is the weakest point in the entire theory of electrons. One imagines that the electron possesses a certain volume and a certain finite charge density, whether due to a volume or surface charge, one cannot manage without that; this, however, conflicts in a certain sense with the atomistic conception of electricity. These are not impossibilities but difficulties, and I am almost surprised that this has not met with more opposition.

This is the point, I believe, at which the quantum theory can be employed with advantage. We can stipulate the laws for large time intervals only. But for small time intervals and great accelerations we still face a gap whose filling requires new hypotheses. Perhaps we may be allowed to assume that an oscillating resonator does not have a continuously variable energy, but that its energy is a simple multiple of an elementary quantum instead. I believe that by using this theory one can arrive at a satisfactory theory of radiation. The question is, then: How does one visualize something like that? That is to say, one asks for a mechanical or electrodynamic model of such a resonator. But mechanics and current electrodynamics do not provide for discrete elements of action, and hence we cannot produce a mechanical or electrodynamic model. Thus, mechanically this seems impossible, and we will have to get used to that. After all, our attempts to mechanically represent the luminiferous ether also have failed completely. There were also attempts to conceive the electric current in a mechanistic way, and to compare it with a stream of water, but this too had to be abandoned, and as one became used to that, so one will have to get used to such a resonator. Of course, this theory would have to be worked out in much greater detail than has been done so far; perhaps someone else will have more luck with it than I had. In any case, I think that first of all one should attempt to transfer the whole problem of the quantum theory to the area of *interaction* between matter and radiation energy; the processes in pure vacuum could then temporarily be explained with the aid of the Maxwell equations.

H. Ziegler: If the *uratoms* of matter are conceived as invisible tiny spheres that possess unchanging speed of light, then it is possible to describe all interactions of corpuscular states and electromagnetic phenomena, and this would also establish the bridge between material and nonmaterial entities still missed by Mr. Planck.

Stark: Mr. Planck pointed out that we have no reason at the moment to switch over to the Einstein consequence, to consider the radiation in space, where it occurs detached from matter, as concentrated. I was originally also of the opinion that for the time being we could restrict ourselves to reducing the elementary law to a certain mode of action of the resonators. But I do believe that there exists a phenomenon that leads to the conclusion that electromagnetic radiation detached from matter, in space, must be considered as concentrated. I have in mind the phenomenon that even at great distances, [2] up to 10 m, electromagnetic radiation that has left an X-ray tube for the surrounding space can still achieve concentrated action on a single electron. I believe that this phenomenon does represent a reason for considering the question of whether the energy of electromagnetic radiation should not be considered as concentrated even where it occurs detached from matter.

Rubens: The view represented by Mr. Einstein would seem to yield a practical conclusion that can be tested experimentally. As we know, it is not only the α-rays, but also the β-rays that produce a scintillating luminous effect on the fluorescent screen. According to the view presented, the same must also hold for γ-rays and X-rays.

Planck: The X-rays are a special case; I would not assert too much about them. - Stark brought up something in favor of the quantum theory, and I wish to bring up something against it; I have in mind the interferences at the enormous phase differences of hundreds of thousands of wavelengths. When a quantum interferes with itself, it would have to have an extension of hundreds of thousands of wavelengths. This is also a certain difficulty.

Stark: The interference phenomena can easily be pitted against the quantum hypothesis. However, once they are treated with more benevolence toward the quantum hypothesis, one will find an explanation for them, too — this is my hope. As for the experimental aspect, it must be emphasized that the experiments to which Mr. Planck alluded involve very dense radiation, so that a very large number of quanta of the same frequency were concentrated in the beam of light; this must be taken into account when discussing those interference phenomena. With radiation of very low density, the interference phenomena would most likely be different. [3]

Einstein: It probably wouldn't be as difficult to incorporate the interference phenomena as one thinks, and the reasons for this are as follows: [4] it must not be assumed that radiations consist of noninteracting quanta; this would make it impossible to explain the phenomena of interference. I picture a quantum as a singularity surrounded by a large vector field. By using a large number of quanta one can construct a vector field that does not differ much from the kind of vector field we assume to be involved in radiations. I can well imagine that when rays impinge upon a boundary surface, a separation of the quanta takes place, due to interaction at the boundary surface, possibly according to the phase of the resulting field at which the quanta reach the interface. The equations for the resulting field would probably not be very different from those in the prevailing theory. It might not be necessary, with respect to interference, to change much in the currently prevailing conceptions. I would like to compare this with the process of molecularization of the carriers of the electrostatic field. The field, as brought about by atomized electric particles, is essentially not very different from the previous conceptions, and it is not out of the question that something similar will happen in the theory of radiation. I do not see any fundamental difficulty in the interference phenomena.

Doc. 62

"Discussion" following lecture version of Fritz Hasenöhrl, "On the Transformation of Kinetic Energy into Radiation" [1]

[*Physikalische Zeitschrift* 10 (1909): 830]

Einstein: Of course, in this case the irregular fluctuations would be exactly such that Maxwell's distribution law would be maintained, i.e., that the damping would be compensated by the irregular impacts. [2]

Lecturer: I am not sure that I understand you correctly. You mean, if you think of a closed container, and particles are moving in it, that they wouldn't be damped by radiation at all?

Einstein: That's right.

Lecturer: I am getting noticeable damping of motion only after a practically infinite time.

Planck: The two gentlemen start out from different assumptions. The lecturer considers an entirely uniform radiation intensity, while Mr. Einstein considers radiation fluctuations, and therefore he also gets fluctuations in the resulting effects, i.e., not a complete damping. [3]