250 BOSTON STUDIES IN
THE PHILOSOPHY OF SCIENCE

The Genesis of General Relativity
Edited by Jürgen Renn

Volume 4

Gravitation in the Twilight of Classical Physics

The Promise of Mathematics

Edited by
Jürgen Renn and Matthias Schemmel

Springer

The Genesis of General Relativity

BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

Editors

ROBERT S. COHEN, *Boston University*
JÜRGEN RENN, *Max Planck Institute for the History of Science*
KOSTAS GAVROGLU, *University of Athens*

Editorial Advisory Board

THOMAS F. GLICK, *Boston University*
ADOLF GRÜNBAUM, *University of Pittsburgh*
SYLVAN S. SCHWEBER, *Brandeis University*
JOHN J. STACHEL, *Boston University*
MARX W. WARTOFSKY†, *(Editor 1960–1997)*

VOLUME 250

The Genesis of General Relativity

Edited by Jürgen Renn

Volume 4

GRAVITATION IN THE TWILIGHT OF CLASSICAL PHYSICS: THE PROMISE OF MATHEMATICS

Editors

Jürgen Renn and Matthias Schemmel
Max Planck Institute for the History of Science, Germany

Associate Editors

Christopher Smeenk
UCLA, U.S.A.

Christopher Martin
Indiana University, U.S.A.

Assistant Editor

Lindy Divarci
Max Planck Institute for the History of Science, Germany

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3999-9 (HB)
ISBN-13 978-1-4020-3999-7 (HB)
ISBN-10 1-4020-4000-8 (e-book)
ISBN-13 978-1-4020-4000-9 (e-book)

As a complete set for the 4 volumes

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved
© 2007 Springer
No part of this work may be reproduced, stored in a retrieval system, or transmitted
in any form or by any means, electronic, mechanical, photocopying, microfilming, recording
or otherwise, without written permission from the Publisher, with the exception
of any material supplied specifically for the purpose of being entered
and executed on a computer system, for exclusive use by the purchaser of the work.

TABLE OF CONTENTS

Volume 4

From an Electromagnetic Theory of Matter to a New Theory of Gravitation

Mie's Theories of Matter and Gravitation 623
Christopher Smeenk and Christopher Martin

Source text 1912–1913: Foundations of a Theory of Matter (Excerpts) . . 633
Gustav Mie

Source text 1914: Remarks Concerning Einstein's Theory of Gravitation 699
Gustav Mie

Source text 1915: The Principle of the Relativity of the Gravitational Potential 729
Gustav Mie

Source text 1913: The Momentum-Energy Law in the Electrodynamics of Gustav Mie 745
Max Born

Including Gravitation in a Unified Theory of Physics

The Origin of Hilbert's Axiomatic Method 759
Leo Corry

Hilbert's Foundation of Physics: From a Theory of Everything to a Constituent of General Relativity 857
Jürgen Renn and John Stachel

Einstein Equations and Hilbert Action: What is Missing on Page 8 of the Proofs for Hilbert's First Communication on the Foundations of Physics? 975
Tilman Sauer

Source text 1915: The Foundations of Physics (Proofs of First Communication) 989
David Hilbert

Source text 1916: The Foundations of Physics
(First Communication) 1003
David Hilbert

Source text 1917: The Foundations of Physics
(Second Communication) 1017
David Hilbert

From Peripheral Mathematics to a New Theory of Gravitation

The Story of Newstein or: Is Gravity just another Pretty Force? 1041
John Stachel

Source text 1877: On the Relation of Non-Euclidean
Geometry to Extension Theory 1079
Hermann Grassmann

Source text 1916: Notion of Parallelism on a General Manifold
and Consequent Geometrical Specification of the Riemannian
Curvature (Excerpts) 1081
Tullio Levi-Civita

Source text 1918: Purely Infinitesimal Geometry (Excerpt) 1089
Hermann Weyl

Source text 1923: The Dynamics of Continuous Media
and the Notion of an Affine Connection on Space-Time 1107
Elie Cartan

Index: Volumes 3 and 4 1131

Volume 3

(parallel volume)

Gravitation in the Twilight of Classical Physics: An Introduction 1
Jürgen Renn and Matthias Schemmel

The Gravitational Force between Mechanics and Electrodynamics

The Third Way to General Relativity: Einstein and Mach in Context 21
Jürgen Renn

Source text 1901: Gravitation 77
Jonathan Zenneck

Source text 1900: Considerations on Gravitation 113
Hendrik A. Lorentz

Source text 1896: Absolute or Relative Motion? 127
Benedict and Immanuel Friedlaender

Source text 1904: On Absolute and Relative Motion 145
August Föppl

An Astronomical Road to a New Theory of Gravitation

The Continuity between Classical and Relativistic Cosmology in the Work of Karl Schwarzschild 155
Matthias Schemmel

Source text 1897: Things at Rest in the Universe 183
Karl Schwarzschild

A New Law of Gravitation Enforced by Special Relativity

Breaking in the 4-Vectors: the Four-Dimensional Movement in Gravitation, 1905–1910. 193
Scott Walter

Source text 1906: On the Dynamics of the Electron (Excerpts) 253
Henri Poincaré

Source text 1908: Mechanics and the Relativity Postulate 273
Hermann Minkowski

Source text 1910: Old and New Questions in Physics (Excerpt) 287
Hendrik A. Lorentz

The Problem of Gravitation as a Challenge for the Minkowski Formalism

The Summit Almost Scaled: Max Abraham as a Pioneer of a Relativistic Theory of Gravitation 305
Jürgen Renn

Source text 1912: On the Theory of Gravitation 331
Max Abraham

Source text 1912: The Free Fall 341
Max Abraham

Source text 1913: A New Theory of Gravitation 347
Max Abraham

Source text 1915: Recent Theories of Gravitation 363
Max Abraham

A Field Theory of Gravitation in the Framework of Special Relativity

Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation 413
John D. Norton

Source text 1912: The Principle of Relativity and Gravitation 489
Gunnar Nordström

Source text 1913: Inertial and Gravitational Mass in Relativistic Mechanics 499
Gunnar Nordström

Source text 1913: On the Theory of Gravitation from the Standpoint of the Principle of Relativity 523
Gunnar Nordström

Source text 1913: On the Present State of the Problem of Gravitation . . . 543
Albert Einstein

From Heretical Mechanics to a New Theory of Relativity

Einstein and Mach's Principle. 569
Julian B. Barbour

Source text 1914: On the Relativity Problem 605
Albert Einstein

Source text 1920: Ether and the Theory of Relativity 613
Albert Einstein

FROM AN ELECTROMAGNETIC THEORY OF MATTER TO A NEW THEORY OF GRAVITATION

CHRISTOPHER SMEENK AND CHRISTOPHER MARTIN

MIE'S THEORIES OF MATTER AND GRAVITATION

Unifying physics by describing a variety of interactions—or even all interactions—within a common framework has long been an alluring goal for physicists. One of the most ambitious attempts at unification was made in the 1910s by Gustav Mie. Mie aimed to derive electromagnetism, gravitation, and aspects of the emerging quantum theory from a single variational principle and a well-chosen world function (Hamiltonian). Mie's main innovation was to consider nonlinear field equations to allow for stable particle-like solutions (now called solitons); furthermore he clarified the use of variational principles in the context of special relativity. The following brief introduction to Mie's work has three main objectives.[1] The first is to explain how Mie's project fit into the contemporary development of the electromagnetic worldview. Part of Mie's project was to develop a relativistic theory of gravitation as a consequence of his generalized electromagnetic theory, and our second goal is to briefly assess this work, which reflects the conceptual resources available for developing a new account of gravitation by analogy with electromagnetism. Finally, Mie was a vocal critic of other approaches to the problem of gravitation. Mie's criticisms of Einstein, in particular, bring out the subtlety and novelty of the ideas that Einstein used to guide his development of general relativity.

In September 1913 Einstein presented a lecture on the current status of the problem of gravitation at the 85th *Naturforscherversammlung* in Vienna. Einstein's lecture and the ensuing heated discussion, both published later that year in the *Physikalische Zeitschrift*, reflect the options available for those who took on the task of developing a new theory of gravitation. The conflict between Newtonian gravitational theory and special relativity provided a strong motivation for developing a new gravitational theory, but it was not clear whether a fairly straightforward modification of Newton's theory based on classical field theory would lead to a successful replacement. Einstein clearly aimed to convince his audience that success would require the more radical step of extending the principle of relativity. For Einstein the development of a new gravitational theory was intricately connected with foundational prob-

1 There are several recent, more comprehensive discussions of Mie's work, which we draw on here: (Kohl 2002; Vizgin 1994; 26–38; Corry 1999, 2004, chaps. 6 and 7). Born (1914) gives an insightful, influential reformulation of Mie's framework, and (Pauli 1921, §64, 188–192 in the English translation) and (Weyl 1918, §25, 206–217 (§26) in the English translation of the fourth edition) both give clear contemporary reviews.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

lems in classical mechanics, and in the Vienna lecture he motivated the need to extend the principle of relativity with an appeal to Mach's analysis of inertia. According to Einstein Mach had accurately identified an "epistemological defect" in classical mechanics, namely the introduction of a distinction between inertial and non-inertial reference frames without an appropriate observational basis.[2] The special theory of relativity had replaced Galilean transformations between reference frames with Lorentz transformations, but the principle of relativity still did not apply to accelerated motion. Extending the principle of relativity to accelerated motion depended on an idea Einstein later called "the most fortunate thought of my life," the principle of equivalence. This idea received many different formulations over the years, but in 1913 Einstein gave one version of this principle as a postulate: his second postulate requires the exact equality of inertial and gravitational mass. He further argued that this equality undermines the ability to observationally distinguish between a state of uniform acceleration and the presence of a gravitational field. The principle of equivalence gave Einstein a valuable link between acceleration and gravitation, tying together the problem of gravitation and the problem of extending the principle of relativity. At the time of the Vienna lecture Einstein was in the midst of an ongoing struggle to clarify the connections among Mach's insight, a generalized principle of relativity, and the formal requirement of general covariance, a struggle that would continue for several more years. Although he also drew heavily on classical field theory in his work, he was convinced that this cluster of ideas would provide the key to a new theory of gravitation.

Gustav Mie's approach to the problem of gravitation stands in sharp contrast to Einstein's. In the discussion following the Vienna lecture, Mie pointedly criticized Einstein's requirement of general covariance and complained that Einstein had overlooked other approaches to gravitation, including his own work and that of Max Abraham.[3] Mie commented that Einstein might have missed his theory of gravitation since it was "tucked away in a work on a comprehensive theory of matter" (CPAE 5, Doc. 18, 1262). This remark aptly characterizes where Mie placed the problem of gravitation conceptually; in Mie's approach the problem of gravitation would be solved as a by-product of an extension of classical field theory. The problem of gravitation was one of the issues, among many, that a "comprehensive theory of matter" would resolve. The pressing issue for Mie was to develop a unified field theory that would succeed where earlier attempts at a reduction of mechanics to electromagnetic theory had failed. By way of contrast with Einstein, Mie's project did not lead out of special

2 Einstein discusses these issues in §4 and §9 of the Vienna lecture (Einstein 1913), as well as in part II of (Einstein 1914), both included in this volume. For a thorough discussion of the role of Machian ideas in Einstein's discovery of general relativity, see "The Third Way to General Relativity" (in vol. 3 of this series).

3 In the published version of the lecture Einstein does briefly mention Abraham's theory only to remark that it fails to satisfy his third postulate, namely the requirement of Lorentz covariance. Mie later noted (Mie 1914, note 13, 175) that the reference to Abraham was only added in the published version of the lecture.

relativity, and Mie was not convinced by Einstein's attempt to link issues in the foundations of mechanics to the problem of gravitation. In Vienna, Einstein justified his sin of omission by pointing out that Mie's theory violated one of his starting assumptions, namely the principle of equivalence. But this clearly did not sway Mie, who expressed doubts that the principle could serve as the basis for a theory and whether it even held in Einstein's own *Entwurf* theory.[4] Mie was also a forceful critic of Einstein's search for a generalized principle of relativity. In the discussion following the Vienna talk and in subsequent articles (Mie 1914, 1915), Mie argued that Einstein had failed to establish a clear link between a principle of general relativity and accelerated motion and questioned the physical content of the principle. Mie had put his finger on the ambiguity of Einstein's guiding principles and the slippage between these ideas and the formal requirement of general covariance. More generally, Mie's criticisms illustrate that Einstein's idiosyncratic path to developing a new gravitational theory seemed to lead into the wilderness in 1913, and that Einstein had not provided entirely convincing reasons to abandon a more conservative path toward a new theory.

Mie's comprehensive theory of matter was presented in a series of three ambitious papers in 1912–13. Mie was eleven years older than Einstein and had held a position as a theoretical physicist in Greifswald since 1902. He was well known for work in applied optics and electromagnetism, including an insightful treatment of the scattering of electromagnetic radiation by spherical particles (Mie 1908) and a widely used textbook (Mie 1910). Mie's textbook endorsed the electromagnetic worldview prominently advocated in the previous decade by Wilhelm Wien and Max Abraham. This worldview amounted to the claim that electromagnetic theory had replaced mechanics as the foundation of physical theory, and Mie characterized electromagnetic theory as "aether physics." Mie emphasized the appeal of reducing physics to a simple set of equations governing the state of the aether and its dynamical evolution, and conceiving of elementary particles as stable "knots" in the aether rather than independent entities (Mie 1912a, 512–13). The aim of the trilogy on matter theory was to develop a unified theory able to account for the existence and properties of electrons (as well as atoms or molecules), explain recent observations of atomic spectra, and yield field equations for gravitation. Although Mie ultimately failed to achieve these grand goals, the approach and formalism he developed influenced later work in unified field theory.

Mie's program differed in important ways from electron theories from the previous decade.[5] The main obstacle to earlier attempts to realize the electromagnetic worldview was the difficulty of explaining the nature and structure of the electron itself in purely electromagnetic terms. Electron theory was an active research area in the first decade of the twentieth century, drawing the attention of many of the best

4 In the discussion, Mie announced that he would soon publish a proof that equality does not hold in the *Entwurf* theory, which appeared in §3 of (Mie 1914).

5 Here we draw primarily on the insightful analysis of the transition from electron theory to relativistic electrodynamics in (Janssen and Mecklenburg 2005); see also the essays collected in (Buchwald and Warwick 2001)

physicists of that generation, such as Lorentz, Abraham, and Sommerfeld. By the time of Mie's work the aim of determining the internal structure of the electron, treated as an extended particle with a definite shape and charge distribution, had been largely abandoned and interest in electron theory had begun to wane. With the advent of special relativity came the realization that the velocity dependence of the electron's mass, a quantity that had been touted as a sensitive experimental test of the internal structure of the electron, was instead a direct consequence of the principle of relativity (Pauli 1921, 185; Pais 1972).[6] Developments in electron theory also threatened the goal of replacing Newtonian mechanics with electromagnetism. Poincaré (1906) proved that an electron treated as a distribution of charge over a spherical shell is not a stable configuration if only the electromagnetic forces are included—the repulsive Coulomb forces would cause it to break apart. Thus it was necessary to introduce the so-called "Poincaré stress," an attractive force needed to maintain the stability of the electron.

One way of responding to these results was to temper the reductive ambitions of the electromagnetic worldview, and to follow Lorentz in admitting charged particles or non-electromagnetic forces as basic elements of the theory. Mie took a different route, and chose instead to alter the field equations of electromagnetism so that there are solutions corresponding to stable particles. A successful theory along these lines would describe the fundamental particles as stable solutions to a set of field equations (with laws of motion derived directly from the field equations) without introducing particles as independent entities, and in this sense reduce mechanics to (generalized) electrodynamics. In effect, Mie treated Maxwell's equations as a weak-field limit of more general field equations. In order to allow for stable charge configurations such as an electron Mie considered non-linear field equations. The fundamental desideratum for the theory was to find generalized field equations that admitted stable solutions representing elementary particles and also reduced to Maxwell's equations in an appropriate limit for regions far from the particles.[7] Mie further aimed to show that gravitation would naturally emerge as a consequence of the generalized field equations.

The key to Mie's theory was the "world function" (Hamiltonian), which he used to derive the field equations via Hamilton's principle.[8] Maxwell's field equations in empty space follow from a Lagrangian $\Phi_{EM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}$ is the Maxwell tensor and the repeated indices (with $\mu,\nu = 1...4$) are summed over. Mie's pro-

6 Lorentz put the point as follows in 1922, "the formula for momentum is a general consequence of the principle of relativity, and a verification of that formula is a verification of the principle and tells us nothing about the nature of mass or of the structure of the electron," quoted in (Janssen and Mecklenburg 2005).

7 Mie was not the first to consider this way of extending classical electromagnetism. Prior to Mie's work Einstein considered replacing Maxwell's field equations with non-linear, inhomogeneous, and/or higher order equations, as reflected in correspondence with Lorentz and Besso in 1908–1910 (see (McCormach 1970) and (Vizgin 1994), 19–26). Einstein, however, was much more keenly aware than Mie of the deep challenges posed by the quantum structure of radiation.

8 Although Mie formulated his theory within a generalized Hamiltonian framework, in the following we focus on the Lagrangian for his field theory (following Born 1914) for ease of exposition.

gram was to find the terms added to Φ_{EM} that would yield the desired generalized field equations. Mie introduced two fundamental assumptions regarding Φ at the outset of the "Grundlagen" (Mie 1912a). First, electrons and other charged particles should be regarded as "states of the aether" rather than independent entities. Mie insisted that the states of the aether should suffice for a complete physical description of matter, although he admitted that failure of his program might force one to enlarge the allowed fundamental variables. Mie distinguished two different types of fundamental variables, the "intensive quantities" and "extensive quantities," treating the latter as analogous to conjugate momenta in Hamiltonian mechanics (see Mie 1915, 254). To enforce the first assumption Mie required that the world function depends only on the field variables (including the electric charge density, the convection current, the magnetic field strength, and the electric displacement). As Born emphasized (1914, 32), this ruled out treating charged particles with trajectories given by independent equations of motion as the source of the field, since including a coupling to a background current in the Lagrangian (i.e., adding a term proportional to $J^{\mu}\phi_{\mu}$) would explicitly introduce dependence on spacetime coordinates. The second assumption was the validity of special relativity, with the consequence that Φ must be Lorentz covariant. The Lagrangian could only include functions of Lorentz invariant terms constructed from $F_{\mu\nu}$ and ϕ_{μ}, the four-vector potential.[9] Mie argued that functions of only two of these invariants, namely $F_{\mu\nu}F^{\mu\nu}$ and $\phi_{\nu}\phi^{\nu}$ should appear in Φ. For the general field equations to reduce to Maxwell's equations, the $\phi_{\nu}\phi^{\nu}$ term could have non-zero values only in regions occupied by particles.

Mie failed to flesh out his formalism with a specific world function satisfying these constraints that led to a reasonable physical theory. Instead he was limited to illustrating his approach with simple examples, such as a Lagrangian $\Phi = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \alpha(\phi_{\nu}\phi^{\nu})^3$, where α is an arbitrary constant. Solutions to the field equations that follow from this Lagrangian could be taken to represent elementary particles, and Mie calculated the charge and mass of the particles. However, these solutions had a number of undesirable features. The arbitrary coefficient appearing in the Lagrangian implied that these solutions placed no constraints on the charge and mass of the "particles," rather than leading to the distinctive values of charge and mass for known particles such as the electron. Mie was further forced to admit (1912b, 38) that his simple world function did not lead to reasonable solutions for interacting charged particles; instead the solutions described a world that eventually separated into two lumps of opposite charge moving away from each other. The simple world

---

9 Since $F_{\mu\nu} = \frac{\partial\phi_{\nu}}{\partial x_{\mu}} - \frac{\partial\phi_{\mu}}{\partial x_{\nu}}$, the Lagrangian depends on ϕ_{ν} and its first derivatives. The list of invariants included the following quantities:

$$\frac{1}{2}F_{\mu\nu};\ \ \phi_{\nu}\phi^{\nu};\ \ F_{\mu\nu}\phi^{\nu}F^{\mu\rho}\phi_{\rho};\ \ (F_{\mu\nu}\phi_{\rho} + F_{\nu\rho}\phi_{\mu} + F_{\rho\mu}\phi_{\nu})^2$$

One invariant was missing from Mie's original list, as Pauli (1921) noted: the quantity $\frac{1}{4}F^{*}_{\mu\nu}\ F^{\mu\nu}$, where $F^{*}_{\mu\nu}$ is the dual of $F_{\mu\nu}$, is an invariant of the restricted Lorentz group, and its square is an invariant of the full Lorentz group.

functions considered by Mie were not viable candidates for a comprehensive description of matter, but he clearly hoped that these problems could be blamed on his lack of ingenuity rather than on his formal framework. However, Pauli (1921, 192) highlighted a problem that went deeper than the failure to find a suitable world function. Mie's world function and the resulting equations of motion both include functions of ϕ_ν. As a result, a stable solution with some value of ϕ is in general not also a solution for ϕ + constant, and the world function also fails to be gauge invariant.[10]

Mie hoped that the appropriate world function (supposing one could be found) would incorporate gravity without needing to put it in by hand. At the outset of the *Grundlagen,* Mie announced his goal of deriving gravity from his matter theory without introducing new dynamical variables and sketched a fanciful picture according to which gravity was a consequence of a cohesive shell or atmosphere binding particles together within an atom (Mie 1912a, 512–514). Mie's description of his project may have raised hopes that the third paper would introduce a truly novel approach to gravitation based on non-linear electrodynamics. But like his other grand goals, this one also eluded Mie's grasp.

Mie's gravitational theory has a great deal in common with competing theories due to Abraham and Nordström. Like Nordström, Mie retained an invariant speed of light and upheld the strict validity of the principle of relativity. This sets his approach apart from Abraham's work; Abraham renounced the constancy of the speed of light and retained the validity of the principle of relativity, restricted to infinitesimal space-time regions, in his first theory, and renounced the principle of relativity all together in his second theory.[11] But like Abraham and Nordström, Mie treated both the source of the gravitational field and the gravitational potential as four-dimensional (Lorentz) scalars, and these were introduced as independent quantities in the world function with no connection to the electromagnetic field. The source of the gravitational field, h, the density of gravitational mass in Mie's theory, is identical to the Hamiltonian density. It is then a short step to derive field equations for the gravitational field appealing to Hamilton's principle. As Mie emphasized, the resulting field equations would be identical to those given by (Abraham 1912) except for the introduction of another variable in the world function.[12] By analogy with his matter theory, Mie introduced an extensive quantity (the excitation of the gravitational field, analogous to electric displacement) conjugate to the gravitational field strength, and argued that the two are identical in an "ideal vacuum" but have an unspecified functional relation in regions occupied by matter.

---

10 Mie recognized this problem and argued that the resulting dependence on the absolute value of the potential would not lead to conflicts with experimental results (Mie 1912b, 24; Mie 1913, 62). Born and Infeld (1934) revived Mie's idea of using a more general Lagrangian, but they excluded additional terms that depended on ϕ_ν to preserve gauge invariance.

11 For discussions of Abraham's and Nordström's theories, see "The Summit Almost Scaled …" and "Einstein, Nordström, and the Early Demise of Scalar, Lorentz Covariant Theories of Gravitation," (both in this volume).

12 See (Mie 1913, 28–29) and the discussion following the Vienna lecture (CPAE 5, Doc. 18).

Mie's failure to achieve a substantial unification illustrates the obstacles to treating gravitation by analogy with electromagnetism. Mie (1915) clearly explained the necessity of introducing the gravitational potential as a dynamical variable in order to resolve the negative energy problem, the most important disanalogy. This problem arises if energy is attributed to the gravitational field itself (as with the electromagnetic field), since the gravitational field strength of, for example, two gravitating masses increases as two masses approach each other, releasing energy in the form of work extracted from the system. One way to save energy conservation in light of this feature of gravitation was to attribute negative energy to the gravitational field, as is suggested by treating Newtonian gravitation in close formal analogy to electrostatics. However, a field with negative energy cannot maintain a stable equilibrium since any small perturbation of the field would in general grow without limit. Following Abraham, Mie argued that the way out of this dilemma was instead to include the gravitational potential in the world function. With this, the internal energy of two approaching masses can be shown to decrease with the decrease of the gravitational potential, thereby compensating for the increase in the field energy.

Including the gravitational potential as a dynamical variable has the consequence that, unlike in electromagnetism, the equations governing physical phenomena depend upon the absolute value of the potential rather than on just potential differences. However, no such dependence had been empirically detected. It remains, moreover, to specify exactly how the field energy depends on the gravitational potential. As (Mie 1915) noted, different choices for this dependence correspond to different gravitational theories. Given the lack of empirical guidance to settle the issue, Mie argued in favor of introducing a principle that would dictate this dependence rather than making what he regarded as arbitrary assumptions. Mie hoped to reconcile his theory's explicit dependence on the absolute value of the gravitational potential with the failure to experimentally detect any such dependence via the theorem (later called a principle) of the relativity of the gravitational potential. The principle plays a central role in the development of Mie's theory, and in elucidating this idea Mie drew a sharp contrast between his approach and Einstein's insistence on generalizing the principle of relativity.

Mie (1915) formulated the principle of the relativity of the gravitational potential as follows:

> In two regions of different gravitational potential exactly the same processes can run according to exactly the same laws if one only thinks of the units of measurement as changing in a suitable way with the value of the gravitational potential. (Mie 1915, 257)

In order to understand the content of this principle, it is perhaps helpful to consider that the principle is equivalent to the requirement that the world function be a homogenous function of the dynamical variables, including the gravitational potential (Mie 1915, 258). From this it immediately follows that in regions of constant gravitational potential, one can transform the potential away, or into any other constant potential, through a rescaling of the remaining dynamical variables and, in general, the spacetime coordinates. Thus, for an observer using correspondingly rescaled measuring

units to measure the dynamical variables, the gravitational potential will be undetectable. Thinking of Mie's principle of relativity of the gravitational potential along these lines as an invariance of the theory under rescaling, we see the gravitational potential and rescalings in Mie's theory as analogous, respectively, to the metric tensor and general linear transformations in Einstein's tensor theory.[13] Simply put, Mie introduces the gravitational potential into the world function to solve the negative energy problem, and introduces an invariance principle, the principle of the relativity of the gravitational potential, to remove any dependence of physical laws on the potential.

By contrast with Einstein, Mie's introduction of this invariance principle for the gravitational field had no connection with foundational problems in mechanics or with extending the principle of relativity. Mie was clearly quite skeptical of the heuristic value of Einstein's guiding ideas. In the discussion following the Vienna lecture, Mie pointedly criticized the idea of extending the principle of relativity to arbitrary states of motion. Mie pressed Einstein to clarify what would be gained by treating a complicated non-uniform motion, such as a bumpy train ride, as physically equivalent to the gravitational field produced by some array of fictitious planets (CPAE 5, Doc. 18). The underlying problem stemmed from Einstein's failure to distinguish between two claims. In the familiar cases of relativity of uniform motion, the two systems in relative motion are entirely physically equivalent. But Einstein's extension of relativity to non-uniform motion involves a very different claim; as he would later clarify, what is relative in the case of non-uniform motion is how the metric field is split into inertial and gravitational components. This does not, however, imply that two observers in non-uniform motion with respect to each other are physically equivalent. In 1913 Einstein did not answer Mie by drawing this distinction; instead, he replied that his theory did not satisfy an entirely general principle of relativity due to a restriction on allowed coordinate transformations (needed, Einstein thought, to insure energy-momentum conservation). Mie (1914) further argued that since the *Entwurf* theory admits only general linear transformations, it does not realize a general principle of relativity, but in fact satisfies precisely Mie's principle of the relativity of the gravitational potential.

Einstein's equivalence principle was also a target of Mie's criticisms. This is not surprising, since Mie's commitment to retaining the framework of special relativity implied that in his theory inertial and gravitational mass would not be exactly equal. Mie (1915, §§5, 6) calculated the effect of the thermal motions of the constituents of bodies on the relation between inertial and gravitational mass, and argued that departures from exact equality would be well within experimental bounds. Exact equivalence could be had at the price of various auxiliary assumptions, according to Mie, but he did not see the need for such extra assumptions, given that his theory fit experimental constraints. He further claimed that Einstein's theory can only guarantee

13 Our understanding here was guided by (CPAE 8, Doc. 346, fn. 3). Note that in his earlier work, Mie (1912, 61) refers to this as the *theorem* of the relativity of the gravitational potential. Even at this early juncture, though, Mie is quick to elevate this theorem, immediately dubbing it a principle.

exact equivalence by making inconsistent assumptions (Mie 1914, 176). This attitude toward the equivalence principle marks another contrast with Einstein, who took the "unity of essence" of inertia and gravitation to be one of the central foundational insights to be respected by his new theory.

In summary, Mie's work illustrates the potential and limitations of approaching the problem of gravitation within the framework of relativistic field theory. Mie's main innovation in the *Grundlagen* was to consider nonlinear field equations, which opened up the possibility of reducing physics to an electromagnetic matter theory. The appeal of this idea has to be balanced against the theory's glaring deficiency, namely the failure to find a particular world function describing even a simple physical system such as two interacting particles. To paraphrase Einstein, although Mie's theory provided a fine formal framework, it was not clear how to fill it with physical content.[14] Even those sympathetic to Mie's program had to admit doubts that this innovation would lead to a successful matter theory, especially given the recent discoveries of quantum phenomena. But whatever the prospects for matter theory based on generalized electrodynamics, Mie's innovations in the *Grundlagen* turned out to provide few insights for developing a gravitational theory. His own gravitational theory shared the insights and limitations of other Lorentz-covariant theories of gravitation.

In terms of the further development of gravitational theories, Mie's influence on David Hilbert is more significant than his own theory. This influence was mediated by Born's clear reformulation of Mie's theory (Born 1914), which showed how Mie's theory fit into the more general framework of (four-dimensional) Lagrangian continuum mechanics as a special case. Mie's project of unification and his mathematical framework, as refined by Born, shaped Hilbert's distinctive path to a new gravitational theory.[15] But this influence depended on Mie's matter theory and not his gravitational theory, which Born and Hilbert both set aside. Furthermore, Hilbert differed from Mie sharply with regard to the status of special relativity. Mie was a persistent critic of Einstein's move to a metric theory of gravitation and saw no reason to leave the framework of special relativity. Hilbert, on the other hand, took Einstein's *Entwurf* theory as one of his starting points, and his synthesis of Mie's matter theory with Einstein's gravitational theory involved replacing the fixed Minkowski metric of special relativity with Einstein's metric tensor. The fertility of Mie's matter theory for Hilbert depended upon setting aside Mie's own gravitational theory as well as his criticisms of Einstein's extension of special relativity.

---

14 In a 1922 letter to Weyl regarding Eddington's later attempt at a unified field theory, quoted in (Vizgin 1994, 37), Einstein commented that "I find the Eddington argument to have this in common with Mie's theory: it is a fine frame, but one cannot see how it can be filled"; see also his negative assessments of Mie's theory (directly or as it was used by Hilbert) in letters to Freundlich (CPAE 5, Doc. 468), Ehrenfest (CPAE 8, Doc. 220) and Weyl (CPAE 8, Doc. 278). Weyl gave a similar assessment of Mie's theory; see §25 of (Weyl 1918); pp. 214–216 of the English translation.

15 This is explored in great detail in "Hilbert's Foundation of Physics …" (in this volume). See also (Sauer 1999) and (Corry 1999, 2004) for assessments of Hilbert's project and the influence of Mie's theory.

REFERENCES

Abraham, Max. 1912. "Relativität und Gravitation. Erwiderung auf eine Bemerkung des Herrn A. Einstein." *Annalen der Physik* (38) 1056–1058.

Born, Max. 1914. "Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie." *Königliche Gesellschaft der Wissenschaften zu Göttingen*. Nachrichten (1914): 23–36. (English translation in this volume.)

Buchwald, Jed Z, and Andrew Warwick (eds.). 2001. *Histories of the Electron. The Birth of Microphysics*. Cambridge: The MIT Press.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.). 1995. *The Collected Papers of Albert Einstein. Vol. 4. The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.). 1996. *The Collected Papers of Albert Einstein. Vol. 6. The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 6E: *The Collected Papers of Albert Einstein. Vol. 6. The Berlin Years: Writings, 1914–1917*. English edition translated by Alfred Engel, consultant Engelbert Schucking. Princeton: Princeton University Press, 1996.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.). 1998. *The Collected Papers of Albert Einstein. Vol. 8. The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press.

CPAE 8E: T*he Collected Papers of Albert Einstein. Vol. 8. The Berlin Years: Correspondence, 1914–1918*. English edition translated by Ann M. Hentschel, consultant Klaus Hentschel. Princeton: Princeton University Press, 1998.

Corry, Leo. 1999. "From Mie's Electromagnetic Theory of Matter to Hilbert's Unified Foundations of Physics." S*tudies in History and Philosophy of Modern Physics* 30 B (2):159–183.

———. 2004. *David Hilbert and the Axiomatization of Physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*. Dordrecht: Kluwer Academic Publishers.

Einstein, Albert. 1913. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14: 1249–1262, (CPAE 4, Doc. 17).

_____.1914. "Zum Relativitätsproblem." *Scientia* 15: 337–348, CPAE 4, Doc. 31.

Janssen, Michel and Mecklenburg, Matthew. Forthcoming. "The Transition from Classical to Relativistic Mechanics: Electromagnetic Models of the Electron." To appear in a volume edited by Jesper Lützen based on the proceedings of the conference, *The Interaction between Mathematics, Physics and Philosophy from 1850 to 1940*, Copenhagen, September 26–28, 2002. Earlier version available as Max Planck Institute for the History of Science Preprint 277.

Kohl, Gunter. 2002. "Relativität in der Schwebe: Die Rolle von Gustav Mie"' Max Planck Institute for the History of Science Preprint 209.

McCormach, Russell. 1970. "Einstein, Lorentz, and the electron theory." *Historical Studies in the Physical and Biological Sciences*, 2: 41–87.

Mie, Gustav. 1908. "Beiträge zur Optik trüber Medien, speziell kolloidaler Metalllösungen," *Annalen der Physik* 25, 378–445.

———. 1910. *Lehrbuch der Elektrizität und des Magnetismus*. Stuttgart: F. Enke.

———. 1914. "Bemerkungen zu der Einsteinschen Gravitationstheorie. I und II." *Physikalische Zeitschrift* 14: 115–122, 169–176." (English translation in this volume.)

———. 1915. "Das Prinzip von der Realität des Gravitationspotentials." In *Arbeiten aus den Gebieten der Physik, Mathematik, Chemie. Festschrift Julius Elster und Hans Geitel zum sechzigsten Geburtstag*. Braunschweig: Friedr. Vieweg & Sohn, 251–268. (English translation in this volume.)

Pais, Abraham. 1972. "The Early History of the Theory of the Electron: 1897–1947." In A. Salam and E. P. Wigner (eds.), *Aspects of Quantum Theory*. Cambridge: Cambridge University Press, 79–93.

Pauli, Wolfgang. 1921. "Relativitätstheorie." In A. Sommerfeld (ed.), *Encyklopädie der mathematischen Wissenschaften, mit Einschluss ihrer Anwendungen*. Leipzig: B. G. Teubner, 539–775.

———. 1958. *Theory of Relativity*. (Translated by G. Field.) London: Pergamon.

Poincaré, Henri. 1906. "Sur la dynamique de l'électron." *Rendiconti del Circolo Matematico di Palermo* 21: 129–175. Reprinted in (Poincaré 1934–54), Vol. 9, 494–550. (English translation of excerpt in vol. 3 of this series.)

Sauer, Tilman. 1999. "The Relativity of Discovery: Hilbert's First Note on the Foundations of Physics." *Archive for History of Exact Science*s 53: 529–575.

Vizgin, Vladimir. 1994. *Unified Field Theories in the First Third of the 20th Century*. (Translated by Barbour, Julian B, edited by E. Hiebert and H. Wussing.) *Science Networks, Historical Studies,* Vol. 13. Basel, Boston, Berlin: Birkhäuser.

Weyl, Hermann. 1918. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. (First edition.) Berlin: Julius Springer.

GUSTAV MIE

FOUNDATIONS OF A THEORY OF MATTER (EXCERPTS)

Originally published as "Grundlagen einer Theorie der Materie" in Annalen der Physik, first communication, 37, 1912, pp. 511–534 and third communication, 40, 1913, pp. 1–65. Greifswald, Physikalisches Institut, 6 January and 31 October 1912. Chapters 2 and 4 are omitted in the translation.

INTRODUCTION

1. The significance of the recently acquired empirical facts about the nature of the atoms ultimately amounts to something essentially only negative, namely that in the atoms' interior the laws of mechanics and Maxwell's equations cannot be valid. But regarding what should replace these equations in order to encompass from a single standpoint the profusion of remarkable facts associated with the notion of quantum of action, and in addition the laws of atomic spectra and so forth, the experimental evidence is silent. In fact, I believe that one must not expect anything like that from experiment alone. Experiment and theory must work hand in hand, and that is not possible as long as the theory has no foundation on which it can be based.

Thus it seems to me absolutely necessary for further progress of our understanding to supply a new foundation for the theory of matter. With this work, I have tried in the following to make a start, but in view of the difficulty of the matter one should not right away expect results accessible to experiment. The immediate goals that I set myself are: to explain the existence of the indivisible electron and: to view the actuality of gravitation as in a necessary connection with the existence of matter. I believe one must start with this, for electric and gravitational effects are surely the most direct expression of those forces upon which rests the very existence of matter. It would be senseless to imagine matter whose | smallest parts did not possess electric [512]
charges, equally senseless however matter without gravitation. Only when the two goals I mentioned are reached will we be able to consider making the connection between the theory and the complex phenomena mentioned above. But achieving both of these goals is still a long way off, and below I can publish only preliminary work, which will perhaps help us to find the way.

The basic assumption of my theory is *that electric and magnetic fields occur also in the interior of electrons*. According to this view, electrons and accordingly the

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

smallest particles of matter in general are not different in nature from the world aether; they are not foreign bodies in the aether, as was thought maybe 20 years ago, but *they are only locations where the aether has taken on a particular state, which we designate by the term electric charge*. However, the enormous intensity of the field- and charge-states at the location itself that we designate as the electron implies that here the usual Maxwell equations are no longer valid. The behavior of the electromagnetic field inside the electron presumably will be very strange when compared to the laws of the "pure aether." But if we can speak at all of an electromagnetic field in the interior of the electron, then it would not be reasonable that there should not be a continuous transition between the behavior of "pure" aether and that of aether in the electron's interior. Therefore, in my theory the electron is not a particle in the aether with a sharp boundary, but consists of a nucleus with a continuous transition into an atmosphere of electric charge that extends to infinity, but which becomes so extraordinarily dilute already quite close to the nucleus that it cannot be experimentally detected in any way. An atom is an agglomeration of a larger number of electrons glued together by a relatively dilute charge of opposite sign. Atoms are probably surrounded by more substantial atmospheres, which however are still so dilute that they
[513] do not cause noticeable | electric fields, but which presumably are asserted in gravitational effects.

It may seem that there is not much to be gained from the basic assumption just formulated. Still, it leads to a general form for the basic equations of the physics of the aether when combined with two further assumptions. The first is *that the principle of relativity shall be valid generally* and the second *that the presently known states of the aether (that is electric field, magnetic field, electric charge, charge current) suffice completely to describe all phenomena of the material world*. The justification of the first assumption should be beyond doubt. Whether the second holds cannot be said *a priori*. An attempt has to be made. If it can produce a theory that mirrors the material world correctly, then it is thereby vindicated. In the opposite case one will have to ask how the system of fundamental quantities is to be enlarged.

In the following I will present in some detail the reasoning that led me from the three assumptions to a general form of the equations of the aether, in order perhaps to stimulate discussion whether the form I assume is the only possible one, or whether there may not also be other basic equations of aether physics consistent with the three assumptions. I admit that I did not succeed in finding other possibilities. That I presuppose the principle of energy conservation as correct, and assume energy to be a localizable quantity, should go without saying.

FIRST CHAPTER: THE FIELD EQUATIONS

General Form of the Field Equations

2. If one considers Maxwell's equations, preferably in the form given to them by Minkowski, one sees immediately that the four-dimensional six-vector "electromag-

netic field strength" in itself does | not suffice to describe the phenomena in space and [514]
time completely. For in addition Maxwell's equations contain an independent four-vector, the "four-current"; which at least therefore has to be included to make the description complete.

In our view, the time component of the four-current, the charge density ρ, represents a peculiar condition of the world aether, which it assumes to a noticeable extent only at isolated places, and which entail that at these places the lines of the electric field $\mathfrak{d}$ simply die out, so that div $\mathfrak{d}$ differs from zero. Therefore we can take the value of div $\mathfrak{d}$ as a measure of this new state of the aether:

$$\rho = \operatorname{div} \mathfrak{d}.$$

Similarly the space component of the four-current, the current of charge $\mathfrak{v}$, describes a peculiar behavior of the aether, which gains noticeable strength only at particular locations; it entails locations of non-vanishing curl of the magnetic field $\mathfrak{h}$ that are not compensated by a time rate of change of the electric field $\mathfrak{d}$. Therefore we can use the difference $\operatorname{curl} \mathfrak{h} - \dot{\mathfrak{d}}$ as a measure of the new state of the aether:

$$\operatorname{curl} \mathfrak{h} - \dot{\mathfrak{d}} = \mathfrak{v}.$$

3. Now we use the basic assumptions mentioned in 1. If all of the material world's processes are to be described by the "electromagnetic field" and the "four-current" together, then by the principle of causality there must be ten differential equations for the ten components of the variables of state $\mathfrak{d}$, $\mathfrak{h}$, ρ, $\mathfrak{v}$ whose left side is always a first-order time derivative of one of the ten quantities, or of a function of them, while on the right-hand side there is a function of the quantities and of their space derivatives. Only such a system of equations can determine from the distribution of the states of the aether at *one* moment the distribution that occurs in the next moment, after passage of an infinitesimal | time dt thus satisfying the principle of causality. [515]

Further, if the principle of relativity is to be valid, then it must be possible to write the derivatives in these equations as vectorial differential operators of four-dimensional quantities. This reduces the number of possibilities enormously. For example, one sees immediately that also with respect to the coordinates only first derivatives can occur; that all derivatives enter only in the first power, etc.

Finally, one must also demand that the equations for the "pure" aether go over into Maxwell's equations, since a continuous transition is assumed between pure aether and matter. Also, the existence of true magnetic charges must be excluded, therefore it must be possible to use a quantity $\mathfrak{b}$ to characterize the magnetic field that everywhere has the property: $\operatorname{div} \mathfrak{b} = 0$. Thus we arrive at the equations:

$$\frac{\partial \mathfrak{d}}{\partial t} = \operatorname{curl} \mathfrak{h} - \mathfrak{v}, \tag{1}$$

$$\frac{\partial \mathfrak{b}}{\partial t} = -\operatorname{curl} \mathfrak{e} \tag{2}$$

and here in pure aether $\mathfrak{b}$ must become identical to $\mathfrak{h}$ and $\mathfrak{e}$ identical to $\mathfrak{d}$ whereas in the interior of matter $\mathfrak{e}$ and $\mathfrak{b}$ can be complicated functions of $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$. Equations (1) and (2) then have only a very superficial resemblance to Maxwell's equations. Since at least half of them are no longer linear, the laws of the field in the interior of atoms are quite different from those in pure aether; for example, electromagnetic waves whose existence presupposes linear equations are excluded there, and further such differences.

Thus in the following we will strictly differentiate between the two "intensive quantities": electric field strength $\mathfrak{e}$, and magnetic induction $\mathfrak{b}$, and the "extensive quantities": electric displacement $\mathfrak{d}$, and magnetic field strength $\mathfrak{h}$. Only in pure aether does the principle of superposition of electromagnetic fields hold, which we will express through the equations $\mathfrak{e} = \mathfrak{d}, \mathfrak{b} = \mathfrak{h}$. |

[516] In terms of symbols of the four-dimensional vector analysis[1] the two equations (1) and (2) take the following form:

$$\Delta t\, v(\mathfrak{h}, -i\mathfrak{d}) = (\mathfrak{v}, i\rho), \tag{1a}$$

$$\Delta t\, v(\mathfrak{e}, i\mathfrak{b}) = 0. \tag{2a}$$

Now the four equations corresponding to the four-vector $(\mathfrak{v}, i\rho)$ must still be dealt with. For a four-vector there are two kinds of four-dimensional first-order differential operator, namely the operators Div and $\mathfrak{Curl}$.[2] In the first operator the time component, and in the second the three space components, are differentiated with respect to t. So we have to use these two operators to obtain the four differential equations that are still missing. The operator Div occurs in the well-known equation:

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathfrak{v} = 0 \tag{3}$$

for this equation becomes in four-dimensional notation:

$$\operatorname{Div}(\mathfrak{v}, i\rho) = 0. \tag{3a}$$

The missing equations must be contained in a formula:

$$\mathfrak{Curl}\,(\mathfrak{f}, i\varphi) = \mathfrak{F}$$

where $(\mathfrak{f}, i\varphi)$ is a four-vector related to $(\mathfrak{v}, i\rho)$ in a similar way as the six-vector $(\mathfrak{b}, -i\mathfrak{e})$ is related to $(\mathfrak{h} - i\mathfrak{d})$. Initially we know only this, that $\mathfrak{f}$ and φ are some functions of all state variables, which taken together form a four-vector. The right side of the equation $\mathfrak{F}$ is some six-vector, also a function of the state variables, of which we know only this, that it must satisfy the condition[3]

---

1 M. Laue, *Das Relativitätsprinzip*, p. 70. Friedr. Vieweg & Sohn, 1911.

2 M. Laue, loc. cit. p. 70.

3 M. Laue, loc. cit. p. 71.

$$\Delta tv\mathfrak{F}^* = 0,$$

because otherwise it could not be obtained from a four-vector by the $\mathfrak{Curl}$ operator. But now, this | condition must further be identical with equation (2a) unless we [517] assume $\mathfrak{F}$ = const., in which case it would admittedly be an identity. For, if this were not the case, then we would have three supernumerary equations, besides the ten differential equations that are necessary for the ten state variables by the principle of causality. The time development of the processes in the aether would then be over-determined, which is of course impossible. Therefore we must necessarily have: either $\mathfrak{F}$ = const. or: $\mathfrak{F} = C \cdot (\mathfrak{b}, -i\mathfrak{e})$, where C is an arbitrary constant factor. We can bring this factor to the other side of our equation $\mathfrak{Curl}(\mathfrak{f}, i\varphi) = \mathfrak{F}$ and absorb it in $\mathfrak{f}, i\varphi$, by simply putting $\mathfrak{F} = (\mathfrak{b}, -i\mathfrak{e})$. The three equations containing a time derivative therefore have this general form:

$$-\frac{\partial \mathfrak{f}}{\partial t} = \nabla\varphi + C \cdot \mathfrak{e} + \mathfrak{c},$$

where C is either zero or one, and $\mathfrak{c}$ denotes a vector that is constant in the entire spacetime region. In a region of pure aether, where $\mathfrak{f} = 0$ as well as $\mathfrak{e} = 0$, it would follow that: $\nabla\varphi = -\mathfrak{c}$. Although all state variables are constant and zero here, φ, which is to be a function of the state variables, would have a non-vanishing gradient, so it would not be constant. This is impossible, therefore we must have: $\mathfrak{c} = 0$. On the other hand it is easy to show that C must be different from zero. If all states of the aether are in equilibrium in the neighborhood of an electron moving with constant velocity, then all time derivatives must be zero. The equation then reads:

$$\nabla\varphi + C \cdot \mathfrak{e} = 0.$$

Now if $C = 0$, then we would also have $\nabla\varphi = 0$, hence φ = const. Then the quantity φ would not depend on the field quantities at all, the same conclusion would hold by the principle of relativity also for $\mathfrak{f}$, and the equation we found would then reduce to an identity. Therefore it must be that $C = 1$. Accordingly, the last three equations of aether dynamics are:

$$-\frac{\partial \mathfrak{f}}{\partial t} = \nabla\varphi + \mathfrak{e}, \tag{4}$$

| or, written in four-dimensional symbols: [518]

$$\mathfrak{Curl}(\mathfrak{f}, i\varphi) = (\mathfrak{b}, -i\mathfrak{e}). \tag{4a}$$

The expression (4a) contains also the following three equations, which contain no time derivative:

$$\mathfrak{curl}\ \mathfrak{f} = \mathfrak{b}. \tag{4b}$$

It is easily seen that the equations (4b) can be derived from (4) with the aid of (2), so they contain nothing new.

If everything is in equilibrium in the vicinity of an electron at rest or in uniform motion, then equation (4) becomes:

$$\nabla\varphi + \mathfrak{e} = 0.$$

We may denote this as the *equilibrium condition* for the field in the vicinity of the electron. It can be interpreted intuitively as saying that the two forces $\mathfrak{e}$ and $\nabla\varphi$ shall be equal and opposite to each other. The electric field strength $\mathfrak{e}$ tends to pull the electron's charge outward and to spread it over as large a region as possible, so it represents the *force of expansion* inherent in matter. It is balanced by the *force of cohesion* $\nabla\varphi$, computed as the gradient of a pressure of cohesion φ peculiar to the electric charge in itself.[4] Forces of expansion and cohesion are the two effects upon which any existence of matter is based, so they must occur in every possible theory of matter.

Equation (4) can be characterized as the equation of motion of the charge current. The vector $\mathfrak{f}$ is the *momentum* [*Bewegungsgröße*] related to the charge current $\mathfrak{v}$. In the usual mechanics we know the momentum to be mass times velocity and measure it by the impact necessary to produce that velocity. Since momentum and pressure are to be characterized as "intensive quantities", that is, quantities to be measured
[519] through | action of forces, we will also contrast φ and $\mathfrak{f}$ as "intensive quantities" with their corresponding "extensive quantities" ρ and $\mathfrak{v}$.

Thus we can describe the state of the aether either by ten extensive quantities $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ or by ten intensive quantities $(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$.

4. The six differential equations (4) and (4b), which are combined in (4a) into one formula, are exactly the same as the differential equations for the so-called four-potential, which is composed of the scalar potential φ and the vector potential $\mathfrak{f}$. Therefore it could be said with some justification that the theory developed here consists simply in attributing to the two potentials φ and $\mathfrak{f}$ the meaning of physical states of the world aether, namely as cohesion pressure and momentum.

Here we must add an important remark. One knows that the solution of equations (4a) for a given six-vector $(\mathfrak{b}, -i\mathfrak{e})$ is undetermined unless one makes a further assumption about $\mathrm{Div}(\mathfrak{f}, i\varphi)$. In the theory of electricity one defines the two potentials by simply putting $\mathrm{Div}(\mathfrak{f}, i\varphi) = 0$. But this equation does not hold for the states of the aether assumed in our theory, and therefore they are in general not identical with the potentials as usually calculated. Namely, the equation of electricity written above is replaced in our aether dynamics by equation (3): $\mathrm{Div}(\mathfrak{v}, i\rho) = 0$. This cannot coexist with the other equation because then the temporal evolution of the aether processes would be governed by eleven equations, and would hence be overdetermined, which is impossible. Therefore in general we have $\mathrm{Div}(\mathfrak{f}, i\varphi) \neq 0$. In a later section (p. 651 [p. 534 in the original]) we will find a simple interpretation for the quantity $\mathrm{Div}(\mathfrak{f}, i\varphi)$.

---

4 As is well known, such a pressure was first assumed by H. Poincaré; (*Compt. rend.* 140, p. 1504, 1905. Cf. also H. Th. Wolff, *Ann. d. Phys.* 36, p. 1066, 1911.

In the case of rest ($\mathfrak{v} = 0$, $\mathfrak{h} = 0$) the quantity φ is indeed identical with the electrostatic potential, because we then have the equation:

$$\mathfrak{e} + \nabla\varphi = 0.$$

| 5. When associating φ with pressure and ρ with a density one could easily [520] believe that it is advantageous always to associate positive values with these quantities, similar to the way it is done in the physics of gases.

We would then ascribe a constant positive value ρ_0, the normal density, to the pure aether when entirely free from fields; for an arbitrary choice of spacetime coordinate system it would of course have to be a four-vector $(\mathfrak{v}_0, i\rho_0)$ that would be constant in the entire spacetime region. Electric and magnetic fields would appear only where ρ and $\mathfrak{v}$ take on values different from ρ_0 and $\mathfrak{v}_0$, and equations (1) and (3) would therefore have to be:

$$\Delta tv(\mathfrak{h}, -i\mathfrak{d}) = ((\mathfrak{v} - \mathfrak{v}_0),\ i\cdot(\rho - \rho_0)),$$
$$\mathrm{Div}((\mathfrak{v} - \mathfrak{v}_0),\ i\cdot(\rho - \rho_0)) = 0.$$

Now, one could of course choose ρ_0 so that the quantity ρ occurring in these equations, the "density of aether", would always be positive. But one cannot see what advantage this would bring. In the following I will therefore always again write simply $\mathfrak{v} - \mathfrak{v}_0$ and $\rho - \rho_0$ instead of $\mathfrak{v}$ and ρ; that is, I will calculate with positive and negative densities by putting the density of pure aether equal to zero.

The same applies to the cohesive pressure φ. Since φ and $\mathfrak{f}$ occur only differentiated with respect to time or space in the fundamental equations of aether physics, one can add to them a quite arbitrary time- and space-independent quantity $\varphi_0, \mathfrak{f}_0$ without changing the description of the processes in any essential way. For example, one could choose a large enough value φ_0, so that $\varphi_0 - \varphi$ always remains positive. The equilibrium condition would then be:

$$\mathfrak{e} - \nabla(\varphi_0 - \varphi) = 0.$$

In pure aether we would now have the large positive pressure φ_0, in an electron we would have the smaller pressure $(\varphi_0 - \varphi)$, and $\mathfrak{e}$ would keep the pressure gradient $-\nabla(\varphi_0 - \varphi)$, which the aether exerts on the electron, in balance. Indeed H. Poincaré (loc. cit.) speaks of a pressure exerted on the electron from the outside. But I believe that for a description it is easier if | we put the zero point of the pressure in the pure [521] aether, and therefore in my calculations I will always put φ equal to zero at infinite distance from an electron.

Similarly for energy, which we know can always be augmented by an arbitrary additive constant, let us fix the zero point in such a way that the energy density in pure, field-free aether equals zero. Similar to ρ and φ the energy density W will then admittedly be allowed to assume negative as well as positive values; but after all there is not the slightest reason that would force us to always make W positive.

With these standardizations, ρ, φ, W are now to be regarded as completely determined quantities without any additive arbitrariness.

The Energy

6. I presuppose that not only the principle of energy conservation, but also the principle of energy localization and energy transfer[5] is valid. That is: if we denote the energy density by W and the energy flux by $\mathfrak{s}$, then from the field equations (1) to (4), the following relation must follow:

$$\frac{\partial W}{\partial t} = -\operatorname{div} \mathfrak{s},$$

where the scalar W as well as the vector $\mathfrak{s}$ are universal functions of the state prevailing at the spacetime point concerned. From the field equations one can arrive at this energy equation in only *one* way: one must determine factors $\mathfrak{k}, \mathfrak{l}, m, \mathfrak{n}$, which are universal functions of the state variables, multiply the equations (1) to (4) by them, and then add the equations. So it must then be possible to pick the factors $\mathfrak{k}, \mathfrak{l}, m, \mathfrak{n}$ in such a way that the left side then becomes a complete time derivative, and the right
[522] side becomes a divergence. | Let us now examine the conditions under which this is possible.

$$\mathfrak{k} \cdot \frac{\partial \mathfrak{d}}{\partial t} + \mathfrak{l} \cdot \frac{\partial \mathfrak{b}}{\partial t} + m \cdot \frac{\partial \rho}{\partial t} + \mathfrak{n} \cdot \frac{\partial \mathfrak{f}}{\partial t}$$
$$= \mathfrak{k} \cdot \operatorname{curl} \mathfrak{h} - \mathfrak{k} \cdot \mathfrak{v} - \mathfrak{l} \cdot \operatorname{curl} \mathfrak{e} - m \cdot \operatorname{div} \mathfrak{v} - \mathfrak{n} \cdot \nabla\varphi - \mathfrak{n} \cdot \mathfrak{e}.$$

First we see that the two terms $-\mathfrak{k} \cdot \mathfrak{v}$ and $-\mathfrak{n} \cdot \mathfrak{e}$, which are pure universal functions of the state variables, must cancel, because $\operatorname{div} \mathfrak{s}$ can consist only of terms containing derivatives with respect to the coordinates. Therefore we must have:

$$\mathfrak{k} = u \cdot \mathfrak{e},$$
$$\mathfrak{n} = -u \cdot \mathfrak{v},$$

where u is again a universal function of the state variables. A small manipulation yields for the right side of the equation:

$$\operatorname{div}(u \cdot [\mathfrak{h} \cdot \mathfrak{e}]) + \operatorname{div}(u \cdot \varphi \cdot \mathfrak{v}) + (u \cdot \mathfrak{h} - \mathfrak{l}) \cdot \operatorname{curl} \mathfrak{e} - \mathfrak{h} \cdot [\mathfrak{e} \cdot \nabla u]$$
$$-(m + u \cdot \varphi) \cdot \operatorname{div} \mathfrak{v} - \varphi \cdot (\mathfrak{v} \cdot \nabla u).$$

This expression can in general be a divergence only if the last four terms are annulled, that is if:

5 G. Mie, *Wiener Sitzungsber.* 107, sec. IIa, p. 1117 and 1126, 1898.

$$\nabla u = 0,$$
$$u \cdot \mathfrak{h} - \mathfrak{l} = 0,$$
$$m + u \cdot \mathfrak{q} = 0.$$

The first of these equations implies $u = \text{const.}$, and specifically the value of this constant is determined by the demand that in pure aether the energy flux should become the well-known Poynting expression $[\mathfrak{d} \cdot \mathfrak{h}] = [\mathfrak{e} \cdot \mathfrak{h}]$. From this it follows that:

$$u = 1, \quad \mathfrak{k} = \mathfrak{e}, \quad \mathfrak{l} = \mathfrak{h}, \quad m = -\mathfrak{q}, \quad \mathfrak{n} = -\mathfrak{v}.$$

Thus we have found for the energy equation:

$$\mathfrak{e} \cdot \frac{\partial \mathfrak{d}}{\partial t} + \mathfrak{h} \cdot \frac{\partial \mathfrak{b}}{\partial t} - \mathfrak{q} \cdot \frac{\partial \rho}{\partial t} - \mathfrak{v} \cdot \frac{\partial \mathfrak{f}}{\partial t} = -\operatorname{div}([\mathfrak{e} \cdot \mathfrak{h}] - \mathfrak{q} \cdot \mathfrak{v}).$$

Accordingly the expression for the *energy flux* in the general aether dynamics is:

$$\mathfrak{s} = [\mathfrak{e} \cdot \mathfrak{h}] - \mathfrak{q} \cdot \mathfrak{v}. \tag{5}$$

| 7. The energy principle further demands that the expression on the left side of the [523]
energy equation be a complete differential. So we must formulate the condition that:

$$\mathfrak{e} \cdot d\mathfrak{d} + \mathfrak{h} \cdot d\mathfrak{b} - \mathfrak{q} \cdot d\rho - \mathfrak{v} \cdot d\mathfrak{f} = dW \tag{6}$$

be a complete differential, so that W can be determined as a function of $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$. Just as well as W we can examine a quantity H, determined by the following equation:

$$W = H + \mathfrak{h} \cdot \mathfrak{b} - \mathfrak{v} \cdot \mathfrak{f}. \tag{7}$$

If W is a function of $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$, then so is H, and vice versa. From (6) and (7) we obtain the following expression for the differential of H:

$$dH = \mathfrak{e} \cdot d\mathfrak{d} - \mathfrak{b} \cdot d\mathfrak{h} - \mathfrak{q} \cdot d\rho + \mathfrak{f} \cdot d\mathfrak{v}, \tag{8}$$

where $\mathfrak{e}, \mathfrak{b}, \mathfrak{q}, \mathfrak{f}$ are functions of $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$. For brevity a vector whose components are

$$\frac{\partial H}{\partial \mathfrak{d}_x}, \quad \frac{\partial H}{\partial \mathfrak{d}_y}, \quad \frac{\partial H}{\partial \mathfrak{d}_z}$$

will be called simply $\partial H / \partial \mathfrak{d}$, and analogously in similar cases. Then it follows directly from (8) that:

$$\mathfrak{e} = \frac{\partial H}{\partial \mathfrak{d}}, \quad \mathfrak{b} = -\frac{\partial H}{\partial \mathfrak{h}}, \quad \mathfrak{q} = -\frac{\partial H}{\partial \rho}, \quad \mathfrak{f} = \frac{\partial H}{\partial \mathfrak{v}}. \tag{9}$$

The condition that the energy principle be valid is that all intensive quantities $\mathfrak{e}, \mathfrak{b}, \mathfrak{q}, \mathfrak{f}$ *can be calculated by means of a single function of the extensive quantities*

$H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$, *which we will call the Hamiltonian function. Specifically, every intensive quantity is to be obtained as the derivative of* H *with respect to the corresponding extensive quantity, in two cases (*$\mathfrak{b}$ *and* φ *) with a negative sign.*

The energy density W can now also be found from the Hamiltonian function alone. For (7) together with (9) result in:

$$W = H - \frac{\partial H}{\partial \mathfrak{h}} \cdot \mathfrak{h} - \frac{\partial H}{\partial \mathfrak{v}} \cdot \mathfrak{v}. \tag{10}$$

The form of the basic equations (1) to (4) of aether dynamics, with equations (9) taken into account, leads immediately to the following theorem: |

[524] *The relativity principle is valid for all physical processes, provided the Hamiltonian function* $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ *is invariant under Lorentz transformations.*

Now we would have the complete formulation of the equations of aether dynamics, if we only knew the form of the universal function H. To find this form is an extremely difficult task indeed.

The problem of a theory of matter is reduced to the problem of finding the universal function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$.

So far we know only one thing about H: in pure aether the superposition principle for electromagnetic fields holds with great precision; so if one takes an additive term $(\mathfrak{d}^2 - \mathfrak{h}^2)/2$ out of H:

$$H = \frac{1}{2}(\mathfrak{d}^2 - \mathfrak{h}^2) + H_1,$$

then the remainder H_1 must be quite vanishingly small compared to the first term at places where ρ is very small. However, in contrast, in the interior of atoms, where ρ is large, H_1 will dominate by far, so that here the laws of the field are quite different than those in pure aether.

8. For calculations it is in general much more convenient to choose the intensive variables $(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$ as the independent variables that describe the state of the aether, and to view the extensive quantities $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ as functions of the former.

Let us now form the following function Φ:

$$\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}) = H - (\mathfrak{e} \cdot \mathfrak{d} - \mathfrak{b} \cdot \mathfrak{h}) + (\varphi \cdot \rho - \mathfrak{f} \cdot \mathfrak{v}), \tag{11}$$

by first solving equations (9) for $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$ as functions of $\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}$ and then substituting the expressions so found on the right side of equation (11). Using (8) we get for the differential of Φ the following expression:

$$d\Phi = -\mathfrak{d} \cdot d\mathfrak{e} + \mathfrak{h} \cdot d\mathfrak{b} + \rho \cdot d\varphi - \mathfrak{v} \cdot d\mathfrak{f}. \tag{12}$$

From this follows:

$$\mathfrak{d} = -\frac{\partial \Phi}{\partial \mathfrak{e}}, \quad \mathfrak{h} = \frac{\partial \Phi}{\partial \mathfrak{b}}, \quad \rho = \frac{\partial \Phi}{\partial \varphi}, \quad \mathfrak{v} = -\frac{\partial \Phi}{\partial \mathfrak{f}}. \tag{13}$$

[525] | *The extensive quantities* $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$ *can all be calculated with the aid of a single function of the intensive quantities* $\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$ *by differentiating the latter with*

respect to the corresponding intensive quantities. In two cases ($\mathfrak{d}$ and $\mathfrak{v}$) one has to give the derivative a negative sign.

The energy density W results from Φ as follows:

$$W = \Phi + \mathfrak{e}\cdot\mathfrak{d} - \varphi\cdot\rho = \Phi - \frac{\partial\Phi}{\partial\mathfrak{e}}\cdot\mathfrak{e} - \frac{\partial\Phi}{\partial\varphi}\cdot\varphi. \tag{14}$$

The Hamiltonian function H is calculated according to (11):

$$H = \Phi - \frac{\partial\Phi}{\partial\mathfrak{e}}\cdot\mathfrak{e} - \frac{\partial\Phi}{\partial\mathfrak{b}}\cdot\mathfrak{b} - \frac{\partial\Phi}{\partial\varphi}\cdot\varphi - \frac{\partial\Phi}{\partial\mathfrak{f}}\cdot\mathfrak{f}. \tag{15}$$

Instead of looking for the universal function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ one can as well search for the universal function $\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f})$.

I will often designate Φ as *the world function* for short.

Under Lorentz transformations Φ as well as H must be an invariant.

Similar to H, one can divide Φ in two parts:

$$\Phi = \frac{1}{2}(\mathfrak{b}^2 - \mathfrak{e}^2) + \Phi_1,$$

the first of which dominates in pure aether, and the second in the interior of atoms.

9. With the aid of the world function one can form a four by four matrix[6] that contains the energy flux and Maxwell's aether stresses for our general aether dynamics:

$$S = \begin{vmatrix} \Phi - \mathfrak{b}\mathfrak{h} + \mathfrak{e}_x\mathfrak{d}_x + \mathfrak{h}_x\mathfrak{b}_x + \mathfrak{f}_x\mathfrak{v}_x, & \mathfrak{e}_x\mathfrak{d}_y + \mathfrak{h}_x\mathfrak{b}_y + \mathfrak{f}_x\mathfrak{v}_y, \\ \mathfrak{e}_x\mathfrak{d}_z + \mathfrak{h}_x\mathfrak{b}_z + \mathfrak{f}_x\mathfrak{v}_z, & -i\cdot(\mathfrak{d}_y\mathfrak{b}_z - \mathfrak{d}_z\mathfrak{b}_y - \rho\mathfrak{f}_x), \\ \mathfrak{e}_y\mathfrak{d}_x + \mathfrak{h}_y\mathfrak{b}_x + \mathfrak{f}_y\cdot\mathfrak{v}_x, & \Phi - \mathfrak{b}\mathfrak{h} + \mathfrak{e}_y\mathfrak{d}_y + \mathfrak{h}_y\mathfrak{b}_y + \mathfrak{f}_y\mathfrak{v}_y, \\ \mathfrak{e}_y\mathfrak{d}_z + \mathfrak{h}_y\mathfrak{b}_z + \mathfrak{f}_y\mathfrak{v}_z, & -i\cdot(\mathfrak{d}_z\mathfrak{b}_x - \mathfrak{d}_x\mathfrak{b}_z - \rho\mathfrak{f}_y), \\ \mathfrak{e}_z\mathfrak{d}_x + \mathfrak{h}_z\mathfrak{b}_x + \mathfrak{f}_z\mathfrak{v}_x, & \mathfrak{e}_z\mathfrak{d}_y + \mathfrak{h}_z\mathfrak{b}_y + \mathfrak{f}_z\mathfrak{v}_y, \\ \Phi - \mathfrak{b}\mathfrak{h} + \mathfrak{e}_z\mathfrak{d}_z + \mathfrak{h}_z\mathfrak{b}_z + \mathfrak{f}_z\mathfrak{v}_z, & -i\cdot(\mathfrak{d}_x\mathfrak{b}_y - \mathfrak{d}_y\mathfrak{b}_x - \rho\cdot\mathfrak{f}_z), \\ -i\cdot(\mathfrak{e}_y\mathfrak{h}_z - \mathfrak{e}_z\mathfrak{h}_y - \varphi\cdot\mathfrak{v}_x), & -i(\mathfrak{e}_z\cdot\mathfrak{h}_x - \mathfrak{e}_x\cdot\mathfrak{h}_z - \varphi\cdot\mathfrak{v}_y), \\ -i\cdot(\mathfrak{e}_x\mathfrak{h}_y - \mathfrak{e}_y\mathfrak{h}_x - \varphi\cdot\mathfrak{v}_z), & \Phi + \mathfrak{e}\mathfrak{d} - \varphi\cdot\rho \end{vmatrix} \tag{16}$$

| If the operation [526]

$$\Delta tv = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} + \frac{\partial}{i\cdot\partial t}$$

is applied to the lowest row of the matrix one obtains the energy equation by putting the expression so obtained equal to zero:

6 H. Minkowski, *Zwei Abhandlungen*. B. G. Teubner 1910, p. 36.

$$\operatorname{div}([\mathfrak{e}\cdot\mathfrak{h}]-\mathfrak{q}\cdot\mathfrak{v})+\frac{\partial}{\partial t}(\Phi+\mathfrak{e}\cdot\mathfrak{d}-\mathfrak{q}\cdot\rho)\;=\;0,$$

because according to (14) we have $\Phi+\mathfrak{e}\cdot\mathfrak{d}-\mathfrak{q}\cdot\rho\;=\;W$. From the principle of relativity it then follows directly that:

$$\Delta tvS\;=\;0. \tag{17}$$

Incidentally, it is also not much effort to obtain the first three rows of S directly from the field equations (1) to (4).

Whether the matrix (16) is symmetric about its diagonal is a question to which we will return again later (p. 650) [p. 533 in original].

Hamilton's Principle

10. Whether the above is an unobjectionable proof that only the form of the field equations as formulated by me is possible, may still be open to discussion. Therefore it seems to me to be valuable to show that the field equations can be obtained by quite simple mathematical operations, assuming the validity of Hamilton's principle.

So I make only the following two assumptions: First, the state of the aether is completely characterized by the quantities $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$, where the last two are defined by the constraints:

$$\rho\;=\;\operatorname{div}\mathfrak{d},\quad \mathfrak{v}\;=\;\operatorname{curl}\mathfrak{h}-\dot{\mathfrak{d}};$$

second, the aether processes satisfy Hamilton's principle formulated as follows.

Hamilton's Principle. There exists a function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$, *whose integral over*
[527] *any spacetime region with determined boundary is an extremum | for all actual processes if the state variables are varied at all points in the interior of the region, but not on the boundary of the region*:

$$\int_G \delta H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})\cdot dx\cdot dy\cdot dz\cdot dt\;=\;0. \tag{18}$$

On the boundary of the region G we have:

$$\delta\mathfrak{d}\;=\;\delta\mathfrak{h}\;=\;\delta\rho\;=\;\delta\mathfrak{v}\;=\;0.$$

It can be shown that the principle of relativity is valid if H is invariant under Lorentz transformations. Let us assume that this is the case and replace the quantities $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$ by the well-known expressions in terms of $\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}'$, that are to take their place upon a transformation of the coordinate system (x, y, z, t) to another (x', y', z', t'); then we must obtain a function $H'(\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}')$, that is built out of the new variables $(\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}')$ in exactly the same way as H is built out of the old variables $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$. We express this by setting: $H'\;=\;H$. G' be the region in the new coordinate system (x', y', z', t') into which G transforms. We then have:

$$\int_{G'} H(\mathfrak{d}', \mathfrak{h}', \rho', \mathfrak{v}') \cdot dx' \cdot dy' \cdot dz' \cdot dt' = \int_{G} H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}) \cdot dx \cdot dy \cdot dz \cdot dt.$$

It follows from this equation that if Hamilton's principle holds for the coordinate system (x, y, z, t), then it also holds for every other arbitrary system (x', y', z', t'). And here the Hamiltonian function is in all coordinate system the same function H.

Accordingly the laws of nature, that is the differential equations resulting from Hamilton's principle, are the same in all coordinate systems that can be obtained by Lorentz transformations. That is the principle of relativity.

Now we want to derive the field equations from Hamilton's principle. To this end we form the variation

$$\delta H = \frac{\partial H}{\partial \mathfrak{d}} \cdot \delta \mathfrak{d} + \frac{\partial H}{\partial \mathfrak{h}} \cdot \delta \mathfrak{h} + \frac{\partial H}{\partial \rho} \cdot \delta \rho + \frac{\partial H}{\partial \mathfrak{v}} \cdot \delta \mathfrak{v}.$$

Now we want to introduce the following abbreviations:

$$\frac{\partial H}{\partial \mathfrak{d}} = \mathfrak{e}, \quad \frac{\partial H}{\partial \mathfrak{h}} = -\mathfrak{b}, \quad \frac{\partial H}{\partial \rho} = -\varphi, \quad \frac{\partial H}{\partial \mathfrak{v}} = \mathfrak{f}. \tag{19}$$

| The variation of H is then: [528]

$$\delta H = \mathfrak{e} \cdot \delta \mathfrak{d} - \mathfrak{b} \cdot \delta \mathfrak{h} - \varphi \cdot \delta \rho + \mathfrak{f} \cdot \delta \mathfrak{v}. \tag{20}$$

To transform this expression further we use a formula from four-dimensional vector calculus, which we want briefly to derive: The product of the four-vector $\mathsf{P} = (\mathfrak{f}, i\varphi)$ and the six-vector $\mathfrak{F} = (\mathfrak{h}, -i\mathfrak{d})$ is taken to be the following four-vector:[7]

$$[\mathsf{P} \cdot \mathfrak{F}] = ([\mathfrak{f} \cdot \mathfrak{h}] + \varphi \cdot \mathfrak{d}, i \cdot (\mathfrak{f} \cdot \mathfrak{d})).$$

We form the Div of this vector:

$$\mathrm{Div}[\mathsf{P} \cdot \mathfrak{F}] = \mathrm{div}\{[\mathfrak{f} \cdot \mathfrak{h}] + \varphi \cdot \mathfrak{d}\} + \frac{\partial(\mathfrak{f} \cdot \mathfrak{d})}{\partial t}.$$

But we have:

$$\mathrm{div}[\mathfrak{f} \cdot \mathfrak{h}] = \mathfrak{h} \cdot \mathrm{curl}\, \mathfrak{f} - \mathfrak{f} \cdot \mathrm{curl}\, \mathfrak{h},$$
$$\mathrm{div}(\varphi \cdot \mathfrak{d}) = \mathfrak{d} \cdot \nabla\varphi + \varphi \cdot \mathrm{div}\, \mathfrak{d},$$
$$\frac{\partial(\mathfrak{f} \cdot \mathfrak{d})}{\partial t} = \mathfrak{d} \cdot \frac{\partial \mathfrak{f}}{\partial t} + \mathfrak{f} \cdot \frac{\partial \mathfrak{d}}{\partial t}.$$

and therefore:

7 M. Laue, *Das Relativitätsprinzip*, p. 67.

$$\operatorname{div}\{[\mathfrak{f}\cdot\mathfrak{h}]+\varphi\cdot\mathfrak{d}\}+\frac{\partial(\mathfrak{f}\cdot\mathfrak{d})}{\partial t} = \mathfrak{h}\cdot\operatorname{curl}\mathfrak{f}+\mathfrak{d}\cdot\left(\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right) - \mathfrak{f}\cdot\left(\operatorname{curl}\mathfrak{h}-\frac{\partial\mathfrak{d}}{\partial t}\right)+\varphi\cdot\operatorname{div}\mathfrak{d}. \tag{21}$$

In four-dimensional symbols this formula becomes:

$$\operatorname{Div}[\mathsf{P}\cdot\mathfrak{F}] = -(\mathfrak{F}\cdot\mathfrak{C}\operatorname{url}\mathsf{P})-(\mathsf{P}\cdot\Delta tv\mathfrak{F}). \tag{22}$$

Now we want to apply this formula to our problem, noting that:

$$\operatorname{curl}\mathfrak{h}-\frac{\partial\mathfrak{d}}{\partial t} = \mathfrak{v},\quad \operatorname{div}\mathfrak{d} = \rho.$$

Therefore we have:

$$\operatorname{Div}[\mathsf{P}\cdot\mathfrak{F}] = \mathfrak{h}\cdot\operatorname{curl}\mathfrak{f}+\mathfrak{d}\cdot\left(\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right)-\mathfrak{f}\cdot\mathfrak{v}+\varphi\cdot\rho.$$

Replacing $\mathfrak{d}$ and $\mathfrak{h}$ by the variation $\delta\mathfrak{d}$ and $\delta\mathfrak{h}$ we find:

$$\mathfrak{f}\cdot\delta\mathfrak{v}-\varphi\cdot\delta\rho = \operatorname{curl}\mathfrak{f}\cdot\delta\mathfrak{h}+\left(\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right)\cdot\delta\mathfrak{d}-\operatorname{Div}[\mathsf{P}\cdot\delta\mathfrak{F}].$$

[529] | Now, the integral:

$$\int_G \operatorname{Div}[\mathsf{P}\cdot\delta\mathfrak{F}]\cdot dx\cdot dy\cdot dz\cdot dt,$$

can, just like the volume integral over a three dimensional divergence, be converted into an integral over the boundary of G. But since Hamilton's principle prescribes that the variations of all state variables, including $\delta\mathfrak{F}$, vanish on the boundary, we have:

$$\int_G \operatorname{Div}[\mathsf{P}\cdot\delta\mathfrak{F}]\cdot dx\cdot dy\cdot dz\cdot dt = 0.$$

Consequently, use of formula (20) for δH results in:

$$\int_G \delta H\cdot dx\cdot dy\cdot dz\cdot dt = \int_G\left(\left(\mathfrak{e}+\nabla\varphi+\frac{\partial\mathfrak{f}}{\partial t}\right)\cdot\delta\mathfrak{d}+(\operatorname{curl}\mathfrak{f}-\mathfrak{b})\cdot\delta\mathfrak{h}\right)\cdot dx\,dy\,dz\,dt.$$

Since there are no longer any constraints between $\mathfrak{d}$ and $\mathfrak{h}$, so that $\delta\mathfrak{h}$ and $\delta\mathfrak{d}$ are quite independent of each other, Hamilton's principle can be satisfied only if the following two differential equations hold:

$$\mathfrak{e} + \nabla\varphi + \frac{\partial \mathfrak{f}}{\partial t} = 0,$$

$$\operatorname{curl} \mathfrak{f} - \mathfrak{b} = 0.$$

These two equations furthermore lead to:

$$\frac{\partial \mathfrak{b}}{\partial t} + \operatorname{curl} \mathfrak{e} = 0.$$

Since the equations of definition (19) for $\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}$ agree completely with equations (9), these equations are identical with the field equations (2) and (4); and equations (1) and (3) we assumed *a priori* as equations of definition.

Hereby it has been proved that the form of the field equations I assumed is the only form in accordance with Hamilton's principle. |

Finally let us remark that equation (21) can be given yet another interesting form [530] by noting that:

$$\operatorname{curl} \mathfrak{f} = \mathfrak{b}, \quad \nabla\varphi + \frac{\partial \mathfrak{f}}{\partial t} = -\mathfrak{e}, \quad \operatorname{curl} \mathfrak{h} - \frac{\partial \mathfrak{d}}{\partial t} = \mathfrak{v}, \quad \operatorname{div} \mathfrak{d} = \rho.$$

Taking into account equation (11) we thus find:

$$\frac{\partial(\mathfrak{f} \cdot \mathfrak{d})}{\partial t} + \operatorname{div}\{[\mathfrak{f} \cdot \mathfrak{h}] + \varphi \cdot \mathfrak{d}\} = \Phi - H. \tag{23}$$

The Invariants

11. In order that the function $H(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ be invariant under Lorentz transformations, i.e. be a four-dimensional scalar, it must be a function of nothing but four-dimensional scalars that can be formed from $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}$. There are four such quantities that are independent of each other.

1. The absolute value of the four-vector $\mathsf{P} = (\mathfrak{v}, i\rho)$. It is:

$$\sigma = \sqrt{\rho^2 - \mathfrak{v}^2} = \rho \cdot \sqrt{1 - \beta^2}, \quad \beta = \frac{\mathfrak{v}}{\rho}.$$

2. The absolute value of the six-vector $\mathfrak{F} = (\mathfrak{h}, -i\mathfrak{d})$. We will take its square:

$$p = \mathfrak{d}^2 - \mathfrak{h}^2.$$

3. The scalar product of the six-vector $\mathfrak{F} = (\mathfrak{h}, -i\mathfrak{d})$ and its dual vector $\mathfrak{F}^* = (-i\mathfrak{d}, \mathfrak{h})$. We will multiply this product by $i/2$ to obtain the quantity

$$q = (\mathfrak{h} \cdot \mathfrak{d}).$$

4. By multiplying the four-vector P by the six-vector $\mathfrak{F}$ and its dual $\mathfrak{F}^*$ one finds two new four-vectors

$$\mathsf{A} = \mathsf{P} \cdot \mathfrak{F} = ((\rho \cdot \mathfrak{d} + [\mathfrak{v} \cdot \mathfrak{h}]), \quad -i \cdot (\mathfrak{v} \cdot \mathfrak{d})),$$
$$\mathsf{B} = \mathsf{P} \cdot \mathfrak{F}^* = (i(\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]), \quad (\mathfrak{v} \cdot \mathfrak{h})).$$

The square of their absolute values are:

$$\mathsf{A}^2 = (\rho \cdot \mathfrak{d} + [\mathfrak{v} \cdot \mathfrak{h}])^2 - (\mathfrak{v} \cdot \mathfrak{d})^2,$$
$$\mathsf{B}^2 = -(\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 + (\mathfrak{v} \cdot \mathfrak{h})^2.$$

[531] | These two quantities are no longer independent of each other, for we can easily see that:

$$\mathsf{A}^2 + \mathsf{B}^2 = (\mathfrak{h}^2 - \mathfrak{d}^2) \cdot (\mathfrak{v}^2 - \rho^2) = \sigma^2 \cdot p.$$

In the same way the scalar product of the two yields nothing new:

$$(\mathsf{A} \cdot \mathsf{B}) = i((\rho \cdot \mathfrak{d} + [\mathfrak{v} \cdot \mathfrak{h}])(\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]) - (\mathfrak{v} \cdot \mathfrak{d}) \cdot (\mathfrak{v} \cdot \mathfrak{h}))$$
$$= -i \cdot (\mathfrak{h} \cdot \mathfrak{d}) \cdot (\mathfrak{v}^2 - \rho^2) = i \cdot \sigma^2 \cdot q.$$

So we get only *one* fourth scalar, for which we will choose the quantity $s = -\mathsf{B}^2$:

$$s = (\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 - (\mathfrak{v} \cdot \mathfrak{h})^2.$$

From the theory of four-dimensional vectors one can prove that there can be no further independent scalars, but I will omit the proof here.

Accordingly we have found as possibilities four independent variables,

$$\begin{cases} \sigma = \sqrt{\rho^2 - \mathfrak{v}^2} = \rho \cdot \sqrt{1 - \dfrac{\mathfrak{v}^2}{\rho^2}}, \\ p = \mathfrak{d}^2 - \mathfrak{h}^2, \\ q = (\mathfrak{d} \cdot \mathfrak{h}), \\ s = (\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 - (\mathfrak{v} \cdot \mathfrak{h})^2. \end{cases} \tag{24}$$

12. The intensive quantities $\mathfrak{e}, \mathfrak{q}, \mathfrak{b}, \mathfrak{f}$ are calculated as follows:

$$\begin{cases} \mathfrak{e} = 2 \cdot \dfrac{\partial H}{\partial p} \cdot \mathfrak{d} + \dfrac{\partial H}{\partial q} \cdot \mathfrak{h} + 2 \cdot \dfrac{\partial H}{\partial s} \cdot [\mathfrak{v} \cdot (\rho \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])], \\ \varphi = -\dfrac{\partial H}{\partial \sigma} \cdot \dfrac{\rho}{\sigma} - 2 \cdot \dfrac{\partial H}{\partial s} \cdot (\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]) \cdot \mathfrak{h}, \\ \mathfrak{b} = 2 \cdot \dfrac{\partial H}{\partial p} \cdot \mathfrak{h} - \dfrac{\partial H}{\partial q} \cdot \mathfrak{d} - 2 \cdot \dfrac{\partial H}{\partial s} \cdot (\rho \cdot (\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]) - \mathfrak{v} \cdot (\mathfrak{v} \cdot \mathfrak{h})), \\ \mathfrak{f} = \dfrac{\partial H}{\partial \sigma} \cdot \dfrac{\mathfrak{v}}{\sigma} - 2 \cdot \dfrac{\partial H}{\partial s} \cdot ([\mathfrak{d} \cdot (\rho \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])] + \mathfrak{h} \cdot (\mathfrak{v} \cdot \mathfrak{h})). \end{cases} \tag{25}$$

By noting that:

$$(\mathfrak{v} \cdot \mathfrak{h}) = \frac{1}{\rho} \cdot (\mathfrak{v} \cdot (\rho \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])),$$

| we recognize immediately that the factor $\partial H / \partial s$ vanishes in the four expressions (25) if: [532]

$$\rho \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}] = 0.$$

If we assume further that $\mathfrak{b} = 0$ in the field of an electron at rest, then $\partial H / \partial q$ must contain either the factor q or the factor s, because otherwise it would not vanish for $\mathfrak{v} = 0$, $\mathfrak{h} = 0$; but now

$$q = (\mathfrak{d} \cdot \mathfrak{h}) = \frac{1}{\rho} \cdot (\mathfrak{d} \cdot (\rho \mathfrak{h} - [\mathfrak{v} \mathfrak{d}])).$$

Accordingly $\partial H / \partial q$ vanishes under the same conditions as the factor $\partial H / \partial s$, namely if:

$$\rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}] = 0.$$

But now one can obtain the quantity $\rho' \cdot \mathfrak{h}' = \rho \cdot \mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]$ by applying a Lorentz transformation to the states of the aether for which one of the coordinate systems moves with respect to the other with a velocity $\mathfrak{q} = \mathfrak{v} / \rho$. If $\mathfrak{q}$ is constant in space and time one can transform to rest, so that $\mathfrak{h}' = 0$, that is: for a stationary motion the condition just written down is satisfied.

If we assume that in the field of an electron at rest not only $\mathfrak{v}$ *and* $\mathfrak{h}$, *but also* $\mathfrak{b}$ *and* $\mathfrak{f}$ *are everywhere zero, then for stationary motion all terms due to the invariants* q *and* s *drop out of the intensive quantities*.

Since all experiences with electrons and matter in general to date refer only to quasistationary motions, and there is no point in burdening the investigations by keeping quantities that presumably will have no influence on the results, we will in the following make the simplifying assumption, that q and s do not occur in H at all.

13. *Hypothesis*. *The Hamiltonian function* H *depends only on the two invariants* σ and p.

Then we have the following very simple expressions for the intensive quantities:

$$\begin{cases} \mathfrak{e} = 2 \cdot \dfrac{\partial H}{\partial p} \cdot \mathfrak{d}, \quad \mathfrak{b} = 2 \cdot \dfrac{\partial H}{\partial p} \cdot \mathfrak{h}, \\ \varphi = -\dfrac{\partial H}{\partial \sigma} \cdot \dfrac{\rho}{\sigma}, \qquad \mathfrak{f} = -\dfrac{\partial H}{\partial \sigma} \cdot \dfrac{\mathfrak{v}}{\sigma}. \end{cases} \tag{26}$$

[533] | *Each of the intensive vectors* $\mathfrak{e}, \mathfrak{b}, \mathfrak{f}$ *is parallel to its corresponding extensive vector* $\mathfrak{d}, \mathfrak{h}, \mathfrak{v}$, *and in addition they are related by the two proportions*:

$$\mathfrak{f} : \mathfrak{v} = \varphi : \rho, \quad \mathfrak{b} : \mathfrak{h} = \mathfrak{e} : \mathfrak{d}.$$

From this follows directly the theorem: *The world matrix (16) is symmetric about its diagonal*.

Like H, so also Φ of course depends only on two variables; for these we will take the following two quantities:

$$\begin{cases} \chi = \sqrt{\varphi^2 - \mathfrak{f}^2}, \\ \eta = \sqrt{\mathfrak{e}^2 - \mathfrak{b}^2}. \end{cases} \tag{27}$$

If we put

$$\frac{\mathfrak{v}}{\rho} = \frac{\mathfrak{f}}{\varphi} = \mathfrak{q},$$

we can also write:

$$\chi = \varphi \cdot \sqrt{1 - \mathfrak{q}^2}. \tag{27a}$$

Finally let us remark that one can find an interesting interpretation for the quantity:

$$\mathrm{Div}(\mathfrak{f}, i\varphi) = \mathrm{div}\, \mathfrak{f} + \frac{\partial \varphi}{\partial t}.$$

I will use the abbreviation:

$$-\frac{1}{\sigma} \cdot \frac{\partial H}{\partial \sigma} = \psi.$$

Then we have:

$$\varphi = \psi \cdot \rho, \quad \mathfrak{f} = \psi \cdot \mathfrak{v},$$

therefore:

$$\mathrm{div}\, \mathfrak{f} + \frac{\partial \varphi}{\partial t} = \psi \cdot \left(\mathrm{div}\, \mathfrak{v} + \frac{\partial \rho}{\partial t} \right) + (\mathfrak{v} \cdot \nabla \psi) + \rho \cdot \frac{\partial \psi}{\partial t}.$$

Now,

$$\mathrm{div}\, \mathfrak{v} + \frac{\partial \rho}{\partial t} = 0$$

and further we can put: $\mathfrak{v} = \rho \cdot \mathfrak{q}$, where we can interpret $\mathfrak{q}$ as the *velocity* with which the charge is being displaced at that place and time. Then we have:

$$(\mathfrak{v} \cdot \nabla\psi) + \rho \cdot \frac{\partial\psi}{\partial t} = \rho \cdot \left(\frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x} \cdot \mathfrak{q}_x + \frac{\partial\psi}{\partial y} \cdot \mathfrak{q}_y + \frac{\partial\psi}{\partial z} \cdot \mathfrak{q}_z \right).$$

Let us think of the several volume elements having charges as individualized, similar to the way we are used to do it with material volume elements, | and consider ψ as a [534] property of the moving element of charge. Then the time rate of change of ψ is:

$$\frac{D\psi}{Dt} = \frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x} \cdot \mathfrak{q}_x + \frac{\partial\psi}{\partial y} \cdot \mathfrak{q}_y + \frac{\partial\psi}{\partial z} \cdot \mathfrak{q}_z.$$

So we arrive at the equation:

$$\operatorname{div} \mathfrak{f} + \frac{\partial\varphi}{\partial t} = \rho \cdot \frac{D\psi}{Dt}. \tag{28}$$

This last equation is of particular interest in view of a theory of gravitation[8] published recently by Abraham. Namely, in a region where the electric field vanishes the quantities that I denote by $\mathfrak{f}_x, \mathfrak{f}_y, \mathfrak{f}_z, i\varphi$ obey the same equations as the quantities called $\mathfrak{F}_x, \mathfrak{F}_y, \mathfrak{F}_z, \mathfrak{F}_u$ by Abraham, with the only difference that Abraham puts:

$$\operatorname{Div} \mathfrak{F} = -4\pi\gamma \cdot \nu,$$

where γ denotes the gravitational constant, ν the mass density, whereas my vector satisfies the equation just derived:

$$\operatorname{Div}(\mathfrak{f}, i\varphi) = \rho \cdot \frac{D\psi}{Dt}.$$

Thus my ansatz would lead to Abraham's theory of gravitation if one wanted to make the assumption that wherever there is material mass, there is a constant increase of the quantity ψ in time. The flux $\mathfrak{f}$ that therefore streams out of the mass particle would be the gravitational field. But since such an assumption is physically absurd, it is excluded to arrive at a gravitational theory in such a simple way from my ansatz. How this probably has to happen has been indicated in the introduction (pp. 633 and 634 [pp. 512 and 513 in the original]).

In the next chapter I will first need to examine whether the existence of indivisible electrons is compatible with my ansatz.

[…]

8 M. Abraham, *Physik. Zeitschr.* 13, p. 1, 1912.

[1]

THIRD CHAPTER: FORCE AND INERTIAL MASS[9]

Calculation of the Force Acting on a Mass Particle

25. To calculate the force we use the world matrix (16), written down in I. p. 525. In doing so we presuppose no restrictions concerning the invariants that enter into the world function, but assume quite generally that all four of the variables (24) enumerated in I. p. 531 occur in H. An easy calculation shows that the theorem established on p. 533 under a restrictive assumption is valid quite generally:

The world matrix is symmetric about its diagonal.

Namely, by applying the multiplication rule

$$[[\mathfrak{a}\cdot\mathfrak{b}]\cdot\mathfrak{c}] = (\mathfrak{a}\cdot\mathfrak{c})\cdot\mathfrak{b} - (\mathfrak{b}\cdot\mathfrak{c})\cdot\mathfrak{a}$$

and the formula that results therefrom

$$[[\mathfrak{a}\cdot\mathfrak{b}]\cdot\mathfrak{c}] + [[\mathfrak{b}\cdot\mathfrak{c}]\cdot\mathfrak{a}] + [[\mathfrak{c}\cdot\mathfrak{a}]\cdot\mathfrak{b}] = 0,$$

one easily finds from the general formula (25) in I. p. 531 the following two equations:

$$[\mathfrak{e}\cdot\mathfrak{d}] + [\mathfrak{h}\cdot\mathfrak{b}] + [\mathfrak{f}\cdot\mathfrak{v}] = 0, \tag{54}$$

$$[\mathfrak{e}\cdot\mathfrak{h}] + [\mathfrak{b}\cdot\mathfrak{d}] + (\rho\cdot\mathfrak{f} - \varphi\cdot\mathfrak{v}) = 0, \tag{55}$$

[2] | and hence, by writing out the components of these expressions,

$$\mathfrak{e}_x\cdot\mathfrak{d}_y + \mathfrak{h}_x\cdot\mathfrak{b}_y + \mathfrak{f}_x\cdot\mathfrak{v}_y = \mathfrak{d}_x\cdot\mathfrak{e}_y + \mathfrak{b}_x\cdot\mathfrak{h}_y + \mathfrak{v}_x\cdot\mathfrak{f}_y \quad \text{etc.},$$
$$\mathfrak{d}_y\cdot\mathfrak{b}_z - \mathfrak{d}_z\cdot\mathfrak{b}_y - \rho\cdot\mathfrak{f}_x = \mathfrak{e}_y\cdot\mathfrak{h}_z - \mathfrak{e}_z\cdot\mathfrak{h}_y - \varphi\cdot\mathfrak{v}_x \quad \text{etc.}$$

The theorem is thereby proved.

26. Let us now imagine a material particle, that is either a location of an electric node or a more complicated structure composed of similar singularities, which moves in an electromagnetic field of large extent. Let s denote the energy flux that is connected with the progressive motion of the states of the aether, as in I. (5) p. 522. Then we have

$$\begin{cases} \mathfrak{s}_x = \mathfrak{e}_y\cdot\mathfrak{h}_z - \mathfrak{e}_z\cdot\mathfrak{h}_y - \varphi\cdot\mathfrak{v}_x = \mathfrak{d}_y\cdot\mathfrak{b}_z - \mathfrak{d}_z\cdot\mathfrak{b}_y - \rho\cdot\mathfrak{f}_x, \\ \mathfrak{s}_y = \mathfrak{e}_z\cdot\mathfrak{h}_x - \mathfrak{e}_x\cdot\mathfrak{h}_z - \varphi\cdot\mathfrak{v}_y = \mathfrak{d}_z\cdot\mathfrak{b}_x - \mathfrak{d}_x\cdot\mathfrak{b}_z - \rho\cdot\mathfrak{f}_y, \\ \mathfrak{s}_z = \mathfrak{e}_x\cdot\mathfrak{h}_y - \mathfrak{e}_y\cdot\mathfrak{h}_x - \varphi\cdot\mathfrak{v}_z = \mathfrak{d}_x\cdot\mathfrak{b}_y - \mathfrak{d}_y\cdot\mathfrak{b}_x - \rho\cdot\mathfrak{f}_z. \end{cases} \tag{56}$$

9 Continuation of the two articles: *Ann. d. Phys.* 37, p. 511, is quoted as I.; *Ann. d. Phys.* 39, p. 1, is cited as II.

We will further define the three-dimensional vectors $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3$ by the following equations:

$$\begin{cases} \Phi - \mathfrak{b}\cdot\mathfrak{h} + \mathfrak{e}_x\cdot\mathfrak{d}_x + \mathfrak{h}_x\cdot\mathfrak{b}_x + \mathfrak{f}_x\cdot\mathfrak{v}_x = \mathfrak{p}_{1x}, \\ \mathfrak{e}_y\cdot\mathfrak{d}_x + \mathfrak{h}_y\cdot\mathfrak{b}_x + \mathfrak{f}_y\cdot\mathfrak{v}_x = \mathfrak{p}_{1y}, \\ \mathfrak{e}_z\cdot\mathfrak{d}_x + \mathfrak{h}_z\cdot\mathfrak{b}_x + \mathfrak{f}_z\cdot\mathfrak{v}_x = \mathfrak{p}_{1z}, \\ \mathfrak{e}_x\cdot\mathfrak{d}_y + \mathfrak{h}_x\cdot\mathfrak{b}_y + \mathfrak{f}_x\cdot\mathfrak{v}_y = \mathfrak{p}_{2x}, \\ \Phi - \mathfrak{b}\cdot\mathfrak{h} + \mathfrak{e}_y\cdot\mathfrak{d}_y + \mathfrak{h}_y\cdot\mathfrak{b}_y + \mathfrak{f}_y\cdot\mathfrak{v}_y = \mathfrak{p}_{2y}, \\ \mathfrak{e}_z\cdot\mathfrak{d}_y + \mathfrak{h}_z\cdot\mathfrak{b}_y + \mathfrak{f}_z\cdot\mathfrak{v}_y = \mathfrak{p}_{2z}, \\ \mathfrak{e}_x\cdot\mathfrak{d}_z + \mathfrak{h}_x\cdot\mathfrak{b}_z + \mathfrak{f}_x\cdot\mathfrak{v}_z = \mathfrak{p}_{3x}, \\ \mathfrak{e}_y\cdot\mathfrak{d}_z + \mathfrak{h}_y\cdot\mathfrak{b}_z + \mathfrak{f}_y\cdot\mathfrak{v}_z = \mathfrak{p}_{3y}, \\ \Phi - \mathfrak{b}\cdot\mathfrak{h} + \mathfrak{e}_z\cdot\mathfrak{d}_z + \mathfrak{h}_z\cdot\mathfrak{b}_z + \mathfrak{f}_z\cdot\mathfrak{v}_z = \mathfrak{p}_{3z}. \end{cases} \tag{57}$$

As we saw in I. on p. 526 eq. (17), the first three rows of the world matrix provide three differential equations, which in consideration of (56) and (57) are to be written as follows:

$$\begin{cases} \dfrac{\partial \mathfrak{s}_x}{\partial t} = \dfrac{\partial \mathfrak{p}_{1x}}{\partial x} + \dfrac{\partial \mathfrak{p}_{2x}}{\partial y} + \dfrac{\partial \mathfrak{p}_{3x}}{\partial z}, \\ \dfrac{\partial \mathfrak{s}_y}{\partial t} = \dfrac{\partial \mathfrak{p}_{1y}}{\partial x} + \dfrac{\partial \mathfrak{p}_{2y}}{\partial y} + \dfrac{\partial \mathfrak{p}_{3y}}{\partial z}, \\ \dfrac{\partial \mathfrak{s}_z}{\partial t} = \dfrac{\partial \mathfrak{p}_{1z}}{\partial x} + \dfrac{\partial \mathfrak{p}_{2z}}{\partial y} + \dfrac{\partial \mathfrak{p}_{3z}}{\partial z}. \end{cases} \tag{58}$$

| Let us now imagine the energy as a fluid, flowing with a certain speed $\mathfrak{q}$. If W is [3] the density of energy, then $\mathfrak{q}$ is determined by the definition

$$\mathfrak{s} = W\cdot\mathfrak{q}. \tag{59}$$

If we further denote by dM the amount of energy occupying at some moment the volume element $dx\cdot dy\cdot dz = dV$, then $dM = W\cdot dV$ and we can as well write equations (58) as follows:

$$\frac{\partial}{\partial t}(dM\cdot\mathfrak{q}_x) = \left(\frac{\partial \mathfrak{p}_{1x}}{\partial x} + \frac{\partial \mathfrak{p}_{2x}}{\partial y} + \frac{\partial \mathfrak{p}_{3x}}{\partial z}\right)\cdot dV,$$

$$\frac{\partial}{\partial t}(dM\cdot\mathfrak{q}_y) = \left(\frac{\partial \mathfrak{p}_{1y}}{\partial x} + \frac{\partial \mathfrak{p}_{2y}}{\partial y} + \frac{\partial \mathfrak{p}_{3y}}{\partial z}\right)\cdot dV,$$

$$\frac{\partial}{\partial t}(dM\cdot\mathfrak{q}_z) = \left(\frac{\partial \mathfrak{p}_{1z}}{\partial x} + \frac{\partial \mathfrak{p}_{2z}}{\partial y} + \frac{\partial \mathfrak{p}_{3z}}{\partial z}\right)\cdot dV.$$

Let us integrate these equations over a volume V. Let

$$M = \int_V dM$$

be the total energy contained in the volume V at the moment we are considering, let $\bar{\mathfrak{q}}$ be the velocity of the "center of mass" in V, defined by the equation:

$$M \cdot \bar{\mathfrak{q}} = \int_V \mathfrak{q} \cdot dM \tag{60}$$

further, let the surface enclosing the volume V be denoted by S, and let N be the outward pointing normal at a point in S; and finally let $\mathfrak{p}_N$ be a three dimensional vector defined by the equation:

$$\mathfrak{p}_N = \mathfrak{p}_1 \cdot \cos(N, x) + \mathfrak{p}_2 \cdot \cos(N, y) + \mathfrak{p}_3 \cdot \cos(N, z) \ldots \tag{61}$$

So the components of $\mathfrak{p}_N$ are computed as follows:

$$\mathfrak{p}_{Nx} = \mathfrak{p}_{1x} \cdot \cos(N, x) + \mathfrak{p}_{2x} \cdot \cos(N, y) + \mathfrak{p}_{3x} \cdot \cos(N, z) \text{ etc.}$$

Integration over V then yields the following result:

$$\frac{\partial (M \cdot \bar{\mathfrak{q}})}{\partial t} = \int_S \mathfrak{p}_N \cdot dS. \tag{62}$$

[4] | Now let us choose the volume V so that it is infinitely small within the extended field in which the material particle moves, but infinitely large in comparison to the enclosed particle. The latter condition is meant to convey, first that the energy of the singularities that constitute the material particle is as good as completely contained in the volume, so that only a quite vanishingly small fraction of the total particle energy resides outside the surface S; and second that on the surface S the vacuum laws already hold as good as exactly, so that ρ and $\mathfrak{v}$ can be taken to be zero, and $\mathfrak{e} = \mathfrak{d}$, $\mathfrak{b} = \mathfrak{h}$. For this choice of the volume V is $M \cdot \bar{\mathfrak{q}}$ the momentum of the particle, its inertial mass is identical with its energy M, and the right side of equations (62) yields the accelerating force [*bewegende Kraft*] acting on the particle. In view of the second condition, and except for vanishingly small correction terms, $\mathfrak{p}_N$ is identical with the component of the Maxwell stress tensor on the corresponding surface element of S; therefore the result for the accelerating force is a value that is independent of the choice of the volume V, provided that the two conditions mentioned above are satisfied; and the value is perfectly identical with what electron theory would yield for a material particle that would be surrounded by the same electric and magnetic field as the particle under consideration. Exactly as in electron theory the accelerating force does not depend on the specific arrangement of the electric charges and the

electric and magnetic dipoles in the interior of the material particle, as long as the particle's own exterior field is the same; in addition it does not depend on the laws of the cohesive forces that hold the particle together, nor on the laws for the electromagnetic field that take the place of Maxwell's equations in the interior of the particle. Exactly the same theorem, which we here have first encountered for the linear motion $\dot{\mathfrak{q}}$ of the particle, can also be shown directly for its rotational motion. Here the moments of inertia are calculated as in the | usual mechanics, by always putting the [5] energy in place of the inertial mass. It is essential for the proof that we have, according to (54), $\mathfrak{p}_{1y} = \mathfrak{p}_{2x}, \mathfrak{p}_{1z} = \mathfrak{p}_{3x}, \mathfrak{p}_{2z} = \mathfrak{p}_{3y}$.

The ponderomotive forces that cause linear or rotational motion in a material particle as a whole are calculated from the electric and magnetic field in which the particle is located according to exactly the same rules as in the usual theory of electricity. The existence of a special four-vector $(\mathfrak{v}, i\mathfrak{p})$ *in the interior of the particle, and the deviation of the laws of the electromagnetic field from Maxwell's equations in the interior of the particle have no perceptible influence on the exterior ponderomotive forces.*

For example, an electron of total charge e, moving with velocity $\mathfrak{q}$ in an electromagnetic field feels the force, according to our theory:

$$\mathfrak{P} = e \cdot (\mathfrak{e} + [\mathfrak{q} \cdot \mathfrak{b}]). \tag{63}$$

This expression agrees exactly with that taken as the basis of electron theory.

By contrast, the effects of forces in the interior of the elementary particles of matter, which may cause delicate changes in the structure of these particles themselves, are something entirely different than the ponderomotive forces of the ordinary theory of relativity. But they cannot be calculated without knowing the world function.

Among the exterior forces acting on the material particle there is also gravity. The theorem just proved implies *that the basic equations of aether dynamics I. (1) to (4), on which we based the theory so far, do not suffice to explain gravity.* Thus the expectation that I expressed at the beginning of my work (I, top of p. 513) has not been fulfilled. In a later chapter we will examine how the basic equations have to be enlarged in order to include gravity as well.

The Inertial Mass of a Material Particle

27. We understand a material particle quite generally to be a small region in the aether where the state | variables take on enormously large values. In the following [6] we will frequently have to evaluate integrals of some state variables over the whole volume of the particle. This is to be understood as a volume whose exterior boundary is sufficiently distant from the center of the particle that the state variables may be treated as infinitely small. Thus, if the outer boundary of the volume is chosen arbitrarily, only such that the particle is "completely" contained by it, as defined here, then this choice cannot have any appreciable effect on the value of the integral.

When saying that a particle is at rest and unchanging we will mean either that all state variables in the volume occupied by the particle are constant, or that the average value of every state variable at every point in the volume is constant when averaged over a time that is infinitely small as far as the experiment is concerned.

For example, let K be the value of a state variable at a point (x, y, z) of the particle. Further let τ be a time that is infinitely small for the experiment. Then the average value we are talking about is

$$\overline{K} = \frac{1}{\tau} \cdot \int_0^\tau K \cdot dt.$$

It is well known[10] that the equations

$$\frac{\overline{\partial K}}{\partial t} = \frac{\partial \overline{K}}{\partial t}, \quad \frac{\overline{\partial K}}{\partial x} = \frac{\partial \overline{K}}{\partial x}, \text{ etc.}$$

are valid here.

Therefore the conditions that the particle be unchanging and at rest are:

$$\frac{\partial \bar{\mathfrak{d}}}{\partial t} = 0, \quad \frac{\partial \bar{\mathfrak{h}}}{\partial t} = 0, \quad \frac{\partial \bar{\rho}}{\partial t} = 0, \quad \frac{\partial \bar{\mathfrak{v}}}{\partial t} = 0, \text{etc.}$$

Now the basic equations I (1) to (4) entail the two relations:

$$\mathfrak{e} \cdot \mathfrak{d} - \varphi \cdot \rho = -\operatorname{div}(\varphi \cdot \mathfrak{d}) - \mathfrak{d} \cdot \frac{\partial \mathfrak{f}}{\partial t},$$

$$\mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} = -\operatorname{div}(\mathfrak{h} \cdot \mathfrak{f}) - \mathfrak{f} \cdot \frac{\partial \mathfrak{d}}{\partial t}.$$

[7] | Inside a particle at rest we therefore have:

$$\overline{\mathfrak{e} \cdot \mathfrak{d}} - \overline{\varphi \cdot \rho} = -\operatorname{div}(\overline{\varphi \cdot \mathfrak{d}}),$$
$$\overline{\mathfrak{b} \cdot \mathfrak{h}} - \overline{\mathfrak{f} \cdot \mathfrak{v}} = -\operatorname{div}(\overline{\mathfrak{h} \cdot \mathfrak{f}}).$$

If we now integrate over a volume that completely encloses the particle, and note that we may set $\overline{\varphi \cdot \mathfrak{d}}$ and $\overline{\mathfrak{h} \cdot \mathfrak{f}}$ equal to zero on the surface of the volume, we find *for a material particle at rest*:

$$\int \overline{\mathfrak{e} \cdot \mathfrak{d}} \cdot dV = \int \overline{\rho \cdot \varphi} \cdot dV, \tag{64}$$

10 H. A. Lorentz, *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegte Körpern,* p.13

$$\int \overline{\mathfrak{b} \cdot \mathfrak{h}} \cdot dV = \int \overline{\mathfrak{f} \cdot \mathfrak{v}} \cdot dV . \tag{65}$$

According to I (7) and (14) on p. 523 and p. 525 the energy density is calculated to be:

$$W = H + \mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} = \Phi + \mathfrak{e} \cdot \mathfrak{d} - \varphi \cdot \rho .$$

So the result for *the energy* E_0 *of a material particle at rest* is according to (64) and (65):

$$E_0 = \int \overline{H} \cdot dV = \int \overline{\Phi} \cdot dV . \tag{66}$$

Let S be some unbounded[11] surface that cuts through to particle, and let N be the surface normal at some point. Since on either side of the surface there must occur no permanent energy changes *provided the particle is at rest*, we must have

$$\int_S \bar{\mathfrak{s}}_N \cdot dS = 0 , \tag{67}$$

where $\bar{\mathfrak{s}}_N$ is the average value of the component of the vector $\mathfrak{s}$ normal to S. According to (56) this vector is given by

$$\mathfrak{s} = [\mathfrak{e} \cdot \mathfrak{h}] - \varphi \cdot \mathfrak{v} = [\mathfrak{d} \cdot \mathfrak{b}] - \rho \cdot \mathfrak{f} .$$

Laue[12] has shown that as long as equation (67) is valid—and this is the case for any material particle—the following theorem is also valid:

Theorem of Laue. *The integral of each component of the world matrix over the volume of a static material | particle is zero, except for only the component with the index 4,4, which yields the energy of the particle.* [8]

In general we here have to take the average of each component over a short time, as in equation (67).

As M. Laue has shown, this theorem can be used to calculate the energy of a moving particle. I will carry out this calculation for the theory being advanced here. Let all the field quantities at a point x_0, y_0, z_0 of the static particle be characterized by the index 0. From these, according to the theory of relativity, one can find the values at a point x, y, z of a moving particle, having speed q in the direction of the z-axis, and if this point (x, y, z) has at time t the position given by the following equations:

$$x = x_0 , \quad y = y_0 , \quad \frac{z - q \cdot t}{\sqrt{1 - q^2}} = z_0$$

according to the following transformation formulas:

11 i.e. either closed or extending to infinity

12 M. Laue, *Das Relativitätsprinzip*, p. 168 ff.

$$\mathfrak{d}_x = \frac{\mathfrak{d}_{x0} + q \cdot \mathfrak{h}_{y0}}{\sqrt{1-q^2}}, \quad \mathfrak{d}_y = \frac{\mathfrak{d}_{y0} - q \cdot \mathfrak{h}_{x0}}{\sqrt{1-q^2}}, \quad \mathfrak{d}_z = \mathfrak{d}_{z0},$$

$$\mathfrak{h}_x = \frac{\mathfrak{h}_{x0} - q \cdot \mathfrak{d}_{y0}}{\sqrt{1-q^2}}, \quad \mathfrak{h}_y = \frac{\mathfrak{h}_{y0} + q \cdot \mathfrak{d}_{x0}}{\sqrt{1-q^2}}, \quad \mathfrak{h}_z = \mathfrak{h}_{z0},$$

$$\rho = \frac{\rho_0 + q \cdot \mathfrak{v}_{z0}}{\sqrt{1-q^2}}, \quad \mathfrak{v}_x = \mathfrak{v}_{x0}, \quad \mathfrak{v}_y = \mathfrak{v}_{y0}, \quad \mathfrak{v}_z = \frac{\mathfrak{v}_{z0} + q \cdot \rho_0}{\sqrt{1-q^2}}.$$

Exactly the same relations as those between $(\mathfrak{d}, \mathfrak{h})$ and $(\mathfrak{d}_0, \mathfrak{h}_0)$ also hold between $(\mathfrak{e}, \mathfrak{h})$ and $(\mathfrak{e}_0, \mathfrak{h}_0)$, and the same as those between $(\rho, \mathfrak{v})$ and $(\rho_0, \mathfrak{v}_0)$ also hold between $(\varphi, \mathfrak{f})$ and $(\varphi_0, \mathfrak{f}_0)$.

The application of these formulas leads through some quite elementary calculations to the following equation:

$$\begin{aligned}\mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} = {} & \mathfrak{b}_0 \cdot \mathfrak{h}_0 - \mathfrak{f}_0 \cdot \mathfrak{v}_0 + \frac{q^2}{1-q^2} \cdot (\mathfrak{e}_0 \mathfrak{d}_0 - \varphi_0 \cdot \rho_0) \\ & - \frac{q^2}{1-q^2} \cdot (\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0} - \mathfrak{b}_{x0} \cdot \mathfrak{h}_{x0} - \mathfrak{b}_{y0} \mathfrak{h}_{y0} + \mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}) \\ & - \frac{q^2}{1-q^2} \cdot ([\mathfrak{e}_0 \cdot \mathfrak{h}_0]_z - \varphi_0 \cdot \mathfrak{v}_{z0} + [\mathfrak{d}_0 \cdot \mathfrak{b}_0]_z - \rho_0 \cdot \mathfrak{f}_{z0}).\end{aligned}$$

Now we form the time average and integrate over the volume occupied by the
[9] material particle. By applying the equations (64), (65), (67) I and noting the relation, in consequence of the definition of the point x, y, z:

$$dx \cdot dy \cdot dz = \sqrt{1-q^2} \cdot dx_0 \cdot dy_0 \cdot dz_0$$

or

$$dV = \sqrt{1-q^2} \cdot dV_0,$$

we reach the result:

$$\left\{ \begin{aligned} & \int (\overline{\mathfrak{b} \cdot \mathfrak{h}} - \overline{\mathfrak{f} \cdot \mathfrak{v}}) \cdot dV = \frac{q^2}{1-q^2} \\ & \qquad \cdot \int (\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} - \overline{\mathfrak{b}_{z0} \cdot \mathfrak{h}_{z0}} - \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV. \end{aligned} \right. \tag{68}$$

If we denote the value of the quantity H at the point x_0, y_0, z_0 of the static particle by H_0, we can regard

$$H_0 = F(x_0, y_0, z_0)$$

as a function of (x_0, y_0, z_0). Further, let (x, y, z) be the point of the moving particle that is obtained at time t by the Lorentz transformation from (x_0, y_0, z_0). Since H is

an invariant under Lorentz transformations, its value at the point (x, y, z) of the moving particle at time t is to be calculated as:

$$H = F\left(x, y, \frac{z - qt}{\sqrt{1-q^2}}\right),$$

where F denotes exactly the same function as above. From this it follows that:

$$\int H \cdot dV = \sqrt{1-q^2} \cdot \int H_0 \cdot dV_0 = \sqrt{1-q^2} \cdot E_0. \tag{69}$$

Now the energy E of the moving particle results from adding (68) and (69):

$$E = \int (\overline{H} + \overline{\mathfrak{b} \cdot \mathfrak{h}} - \overline{\mathfrak{f} \cdot \mathfrak{v}}) \cdot dV,$$

$$E = \sqrt{1-q^2} \cdot \int \overline{H}_0 \cdot dV_0 + \frac{q^2}{\sqrt{1-q^2}} \cdot \int (\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} - \overline{\mathfrak{v}_{z0} \cdot \mathfrak{h}_{z0}} - \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV_0.$$

This result can be simplified further with the aid of Laue's theorem. Namely, if we apply this theorem | to the term of the world matrix (16) with index 3,3 we get: [10]

$$\int (\overline{\Phi}_0 - \overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} + \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} + \overline{\mathfrak{h}_{z0} \cdot \mathfrak{b}_{z0}} + \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV_0 = 0.$$

Thus, since according to (66):

$$\int \overline{\Phi}_0 \cdot dV_0 = \int \overline{H}_0 \cdot dV_0 = E_0,$$

$$\int (\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{e}_{z0} \cdot \mathfrak{d}_{z0}} - \overline{\mathfrak{h}_{z0} \cdot \mathfrak{b}_{z0}} - \overline{\mathfrak{f}_{z0} \cdot \mathfrak{v}_{z0}}) \cdot dV_0 = \int \overline{\Phi}_0 \cdot dV_0 = E_0.$$

The result is what M. Laue has already shown in general (*Das Relativitätsprinzip* p. 170):

$$E = \frac{E_0}{\sqrt{1-q^2}}. \tag{70}$$

28. Another interesting consequence can be derived from Laue's theorem. By applying it to the three terms of the diagonal of the world matrix with the indices 1,1, as well as 2,2 and 3,3 one obtains:

$$\int (\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{x0} \cdot \mathfrak{h}_{x0}} - \overline{\mathfrak{e}_{x0} \cdot \mathfrak{d}_{x0}} - \overline{\mathfrak{f}_{x0} \cdot \mathfrak{v}_{x0}}) \cdot dV_0 = E_0,$$

$$\int (\overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{y0} \cdot \mathfrak{h}_{y0}} - \overline{\mathfrak{e}_{y0} \cdot \mathfrak{d}_{y0}} - \overline{\mathfrak{f}_{y0} \cdot \mathfrak{v}_{y0}}) \cdot dV_0 = E_0,$$

$$\int(\overline{\mathfrak{b}_0\cdot\mathfrak{h}_0}-\overline{\mathfrak{b}_{z0}\cdot\mathfrak{h}_{z0}}-\overline{\mathfrak{e}_{z0}\cdot\mathfrak{d}_{z0}}-\overline{\mathfrak{f}_{z0}\cdot\mathfrak{v}_{z0}})\cdot dV_0 \;=\; E_0.$$

By addition of these equations one gets:

$$\int(2\cdot\overline{\mathfrak{b}_0\mathfrak{h}_0}-\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0}-\overline{\mathfrak{f}_0\cdot\mathfrak{v}_0})\cdot dV_0 \;=\; 3\cdot E_0,$$

or, taking into account (64) and (65):

$$\begin{cases} E_0 \;=\; -\dfrac{1}{3}\cdot\int(\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0}-\overline{\mathfrak{b}_0\cdot\mathfrak{h}_0})\cdot dV_0, \\ \qquad\; =\; -\dfrac{1}{3}\cdot\int(\overline{\rho_0\cdot\varphi_0}-\overline{\mathfrak{f}_0\cdot\mathfrak{v}_0})\cdot dV_0. \end{cases} \tag{71}$$

In addition it is immediately seen from the three equations just written down that:

$$\begin{cases} \quad\int(\overline{\mathfrak{e}_{x0}\cdot\mathfrak{d}_{x0}}+\overline{\mathfrak{b}_{x0}\cdot\mathfrak{h}_{x0}}+\overline{\mathfrak{f}_{x0}\cdot\mathfrak{v}_{x0}})\cdot dV_0 \\ =\int(\overline{\mathfrak{e}_{y0}\cdot\mathfrak{d}_{y0}}+\overline{\mathfrak{b}_{y0}\cdot\mathfrak{h}_{y0}}+\overline{\mathfrak{f}_{y0}\cdot\mathfrak{v}_{y0}})\cdot dV_0 \\ =\int(\overline{\mathfrak{e}_{z0}\cdot\mathfrak{d}_{z0}}+\overline{\mathfrak{b}_{z0}\cdot\mathfrak{h}_{z0}}+\overline{\mathfrak{f}_{z0}\cdot\mathfrak{v}_{z0}})\cdot dV_0 \\ =\dfrac{1}{3}\int(\overline{\mathfrak{e}_0\cdot\mathfrak{d}_0}+\overline{\mathfrak{b}_0\cdot\mathfrak{h}_0}+\overline{\mathfrak{f}_0\cdot\mathfrak{v}_0})\cdot dV_0. \end{cases} \tag{72}$$

[11] | These equations become particularly interesting when $\mathfrak{h}=0, \mathfrak{v}=0$, as is the case for an electron.

In the field of an electron we have:

$$E_0 \;=\; -\frac{1}{3}\cdot\int\mathfrak{e}_0\cdot\mathfrak{d}_0\cdot dV \;=\; -\frac{1}{3}\cdot\int\varphi_0\cdot\rho_0\cdot dV_0 \tag{73}$$

and besides:

$$\begin{cases} \int\mathfrak{e}_{x0}\cdot\mathfrak{d}_{x0}\cdot dV_0 \;=\int\mathfrak{e}_{y0}\cdot\mathfrak{d}_{y0}\cdot dV_0 \;=\int\mathfrak{e}_{z0}\cdot\mathfrak{d}_{z0}\cdot dV_0 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \;\dfrac{1}{3}\int\mathfrak{e}_0\cdot\mathfrak{d}_0\cdot dV_0. \end{cases} \tag{74}$$

29. For the special case discussed thoroughly in II. on pp. 18 ff. the relation (73) can easily be verified. When we substitute into the world function

$$\Phi \;=\; -\frac{1}{2}\eta^2-\frac{1}{6}a\cdot\chi^6$$

the values for the static field, $\eta = \mathfrak{e}_0, \chi = \varphi_0$, the result is:

$$E_0 = \int \Phi_0 \cdot dV_0 = -\frac{1}{2} \cdot \int \mathfrak{e}_0^2 \cdot dV_0 + \frac{1}{6} a \cdot \int \varphi_0^6 \cdot dV_0.$$

But furthermore:

$$\mathfrak{d}_0 = -\frac{\partial \Phi_0}{\partial \mathfrak{e}_0} = \mathfrak{e}_0, \quad \rho_0 = \frac{\partial \Phi_0}{\partial \varphi_0} = a \cdot \varphi_0^5$$

and so we may write:

$$E_0 = -\frac{1}{2} \cdot \int \mathfrak{e}_0 \cdot \mathfrak{d}_0 \cdot dV + \frac{1}{6} \cdot \int \varphi_0 \cdot \rho_0 \cdot dV.$$

When (64) is applied to this, the result is (73).

Had we given the wave function the more general form:

$$\Phi = -\frac{1}{2}\eta^2 + \frac{1}{\nu} \cdot a \cdot \chi^\nu,$$

quite an analogous calculation would yield:

$$E_0 = -\frac{1}{2} \cdot \int \mathfrak{e}_0 \cdot \mathfrak{d}_0 \cdot dV + \frac{1}{\nu} \cdot \int \varphi_0 \cdot \rho_0 \cdot dV,$$

so that relation (73) could not possibly be satisfied except when $\nu = 6$. This implies:

For all wave functions of the form:

$$\Phi = -\frac{1}{2}\eta^2 + \frac{1}{\nu} \cdot a \cdot \chi^\nu$$

| *only the case* $\nu = 6$ *can lead to isolated nodes of electric charge*. [12]

If one takes some different value for ν, then all integrals of equation (34) in II. p. 15 must have essential singularities, either a singularity at the origin, or at infinity, or both. Then there is no single integral that could represent an electron.

From this one sees that equation (73) can be used on occasion as a criterion whether or not a particular form of the wave function is consistent with the existence of isolated nodes (electrons).

30. From formula (73) it follows that the energy of a node is *negative* in the example discussed in II. So in this case the negative energy attributed to the cohesive effect of the charges exceeds the positive energy of the electric field. Since in the Hamiltonian function:

$$H(\mathfrak{d}_0, 0, \rho_0, 0) = \Phi_0 + \mathfrak{e}_0 \mathfrak{d}_0 - \varphi_0 \cdot \rho_0 = W$$

$$= \frac{1}{2}\mathfrak{d}^{02} - \frac{5}{6} \sqrt[5]{\frac{\rho_0^6}{a}}$$

$\mathfrak{d}_0$ and ρ_0 occur quite separate from each other, the two amounts of energy can also be calculated separately. For the energy of the electric field one gets:

$$\frac{1}{2}\cdot\int \mathfrak{d}_0^2\cdot dV_0 \;=\; \frac{1}{2}\cdot\int \mathfrak{e}_0\cdot\mathfrak{d}_0 dV_0,$$

and for the energy of the cohesive forces:

$$-\frac{5}{6}\cdot\int \sqrt[5]{\frac{\rho_0^6}{a}}\cdot dV_0 = -\frac{5}{6}\cdot\int \rho_0\cdot\varphi_0\cdot dV_0 \;=\; -\frac{5}{6}\cdot\int \mathfrak{e}_0\cdot\mathfrak{d}_0\cdot dV_0.$$

But if now the energy of a particle is negative, the same must be true for its inertial mass. The nodes mentioned in II. on p. 37 thus have a negative inertial mass; in fields of force they must accordingly assume accelerations that are exactly opposite to the accelerating forces. This explains the behavior that at first seems absurd, to which we were led in II. p. 38 by general reasoning, namely that equal nodes tend to
[13] congregate, and opposite | nodes tend to separate, although the ponderomotive forces of the electric fields act in precisely the opposite direction.

Another general conclusion can be drawn from (73):

The necessary and sufficient condition that the inertial mass of an electron be positive is:

$$\int \mathfrak{e}_0\cdot\mathfrak{d}_0\cdot dV < 0$$

or equally well:

$$\int \varphi_0\cdot\rho_0\cdot dV < 0.$$

At a large distance from the electron we have $\mathfrak{e} = \mathfrak{d}$, so that $\mathfrak{e}_0\cdot\mathfrak{d}_0$ is certainly positive. This implies:

In the interior of the electron the two vectors $\mathfrak{e}$ *and* $\mathfrak{d}$ *must have opposite sign*.

It is seen from this that it is quite impossible that Maxwell's equations continue to be valid in the interior of an electron.

Similarly φ being equal to the electric potential, has the same sign as ρ in the outer spheres of the electron. In particular, φ reaches its maximum at the place where $\mathfrak{e}$ crosses zero as it assumes the opposite direction in the interior of the electron. Farther in the interior φ must then decrease sufficiently also eventually to change its sign and make $\varphi_0\cdot\rho_0$ so large and negative that the volume integral of $\varphi_0\cdot\rho_0$ must be negative.

In the very interior of the electron φ *must attain the opposite sign to* ρ.

[…]

FIFTH CHAPTER: GRAVITATION [25]

The Extended Basic Equation of the Dynamics of the Aether

37. We saw on p. 655 [p. 5 in the original] that the assumed cohesive pressure of electric charges together with the electromagnetic field still is not sufficient to explain all actions of force in the world of matter. Gravitation is missing, and we are now forced to enlarge the system of fundamental quantities, into which at first we accepted as few quantities as at all possible (I, p. 634 [p. 513 in the original]), namely only the six-vector $(\mathfrak{h}, -i \cdot \mathfrak{d})$ and the four-vector $(\mathfrak{v}, i\rho)$.

It would be most straightforward to conceive of gravity as a cohesive action that resides in the energy itself. But if we want to maintain the validity of the principle of relativity, we cannot allow energy by itself to enter into the extended basic equations, for in relativity theory the energy density is the last entry of the world matrix (cf. I, p. 643, equation (16) [p. 525 in the original]), so the whole matrix as such would have to appear in the equations. One runs into insuperable difficulties if one tries to connect this matrix with some other four-dimensional quantity by means of four-dimensional differential operators, and thus to obtain equations that obey both the causality principle (I. p. 635 [p. 514 in the original]) and the energy principle (I. p. 640 [p. 521 in the original]). | I have struggled for a long time with such attempts, [26] which always led to quite cumbersome systems of equations, and I am convinced that it is quite impossible to attain in this way a theory of gravitation that obeys both the relativity principle and the energy principle.

By contrast it is extraordinarily easy and simple to reach the goal if the cohesive tendency is ascribed not to the quantity W, but to the quantity H, which is defined as $H = W - \mathfrak{b} \cdot \mathfrak{h} + \mathfrak{v} \cdot \mathfrak{f}$ by the equation (7) in I. on p. 641 [p. 523 in the original]. As long as the velocities of the material elementary particles are small compared to the speed of light, it will be experimentally undecidable whether W or H control the gravitational effects. To wit, according to equations (69) and (70) we have for a moving massive particle:

$$\int H dV = \sqrt{1-q^2} E_0,$$

$$\int W dV = \frac{1}{\sqrt{1-q^2}} E_0,$$

where the integrals are to be extended over the volume occupied by the particle, and where E_0 denotes the energy of the particle when at rest, and q the ratio of its velocity to the speed of light. So we see that practically there is no appreciable difference between the two integrals.

But the quantity H is a four-dimensional scalar, and to it the differential operator can be applied in only a single way; this produces a four-vector, the gradient of the scalar. Conversely the four-vector can also be associated with a scalar by applying to

it the "divergence" operation. By contrast a six-vector cannot be related to a scalar through a four-dimensional differential operation of first order. This implies:

The gravitational field must necessarily be represented by a four-vector, not by a six-vector.

This theorem is however based on the supposition that the gravitating mass is to
[27] be numerically represented by a four-dimensional scalar, | namely the quantity H. It would be different if the density of the gravitating mass were the fourth component of a four-vector, such as the density of electric charge. Then the gravitational field would require a six-vector, similar to the electromagnetic field. But as far as I can see it is impossible to find a four-vector whose fourth component would approximately equal the energy density, and the theories of gravitation that treat the gravitational field in the same way as the electromagnetic field, such as those of O. Heaviside,[13] H. A. Lorentz,[14] R. Gans,[15] therefore either cannot be in accord with the principle of relativity, or the gravitational mass cannot be equal to the inertial mass in these theories.

To establish the equations of the gravitational field we proceed in the same way as we did when setting up the basic electromagnetic equation in I, sections 2. to 5. We assume that for a complete description of the material world we need, in addition to the six-vector $(\mathfrak{h}, -i\mathfrak{d})$ and the four-vector $(\mathfrak{v}, i\rho)$, yet another four-vector $(\mathfrak{g}, iu)$ and a scalar ω. This system of quantities is paralleled by a second, which is completely determined if all the quantities of the first system are given. Of the second system we already know the six-vector $(\mathfrak{b}, -i\mathfrak{e})$ and the four-vector $(\mathfrak{f}, i\varphi)$, which however now depend not only on $(\mathfrak{h}, -i\mathfrak{d})$ and $(\mathfrak{v}, i\rho)$ but also on $(\mathfrak{g}, iu)$ and ω. To this we must further add a four-vector $(\mathfrak{k}, iw)$ and a scalar H, which correspond to $(\mathfrak{g}, iu)$ and ω. The scalar H shall be essentially identical to the quantity defined in I, p. 523. However, like the energy density W, it depends not only on $(\mathfrak{h}, -i\mathfrak{d})$ and $(\mathfrak{v}, i\rho)$ but also on the quantities of the gravitational field, that is $(\mathfrak{g}, iu)$ and ω, and the relation (7) will accordingly have to be subjected to a minor alteration. Now we apply one of the two possible four-dimensional vector operations to $(\mathfrak{g}, iu)$ and ω, and we apply the other operation to $(\mathfrak{k}, iw)$ and H. In this way we obtain the only
[28] possible form for the | laws of gravitation that is in accord with the principle of relativity:

13 O. Heaviside, *Electromagnetic Theory* 1, p. 455, 1894.

14 H. A. Lorentz, *Versl. Kon. Ak. Wet.* Amsterdam 8. p. 603, 1900.

15 R. Gans, *Physik. Zeitschr.* 6. p. 803, 1905.

$$\left.\begin{aligned} \mathfrak{g}_x &= \frac{\partial \omega}{\partial x}, \\ \mathfrak{g}_y &= \frac{\partial \omega}{\partial y}, \\ \mathfrak{g}_z &= \frac{\partial \omega}{\partial z}, \\ u &= -\frac{\partial \omega}{\partial t}, \end{aligned}\right\} \tag{85}$$

$$\frac{\partial \mathfrak{k}_x}{\partial x} + \frac{\partial \mathfrak{k}_y}{\partial y} + \frac{\partial \mathfrak{k}_z}{\partial z} + \frac{\partial w}{\partial t} = -\gamma H. \tag{86}$$

Here γ shall denote a universal constant. The equations (85) are equivalent to the following:

$$\left.\begin{aligned} \frac{\partial \mathfrak{g}_x}{\partial t} + \frac{\partial u}{\partial x} &= 0, \\ \frac{\partial \mathfrak{g}_y}{\partial t} + \frac{\partial u}{\partial y} &= 0, \\ \frac{\partial \mathfrak{g}_z}{\partial t} + \frac{\partial u}{\partial z} &= 0, \end{aligned}\right\} \tag{87}$$

$$\frac{\partial \omega}{\partial t} = -u. \tag{88}$$

Equations (86), (87), and (88) together form a system of five mutually independent equations, each containing a first derivative with respect to time of one of the five new state variables. Consequently the causality principle is satisfied.

The complete system of basic equations of the physics of the aether, including the effects of gravitation [Gravitationswirkungen], is given by the equations: (1), (2), (3), (4), (86), (87), (88).

In terms of the symbols of four-dimensional vector analysis the equations (85) to (88) can also be written as follows:

$$\begin{aligned} (\mathfrak{g}, iu) &= \Gamma\rho\alpha\delta\ \omega, \\ \mathrm{Div}(\mathfrak{k}, iw) &= -\gamma \cdot H, \\ \mathfrak{Curl}(\mathfrak{g}, iu) &= 0. \end{aligned}$$

| The system of equations (85) and (86) would formally agree with that upon [29] which M. Abraham[16] bases his theory of gravitation, if one would put the two vectors $(\mathfrak{g}, iu)$ and $(\mathfrak{k}, iw)$ equal to each other. M. Abraham proceeds in his theory from the

16 M. Abraham, *Physik. Zeitschr.* 13. p. 1. 1912.

presupposition that the density of gravitating mass (which he calls ν) is a four-dimensional scalar, and since moreover he makes use of the theory of relativity in the cited paper, he had to arrive at this system of equations, the only one that the theory of relativity can produce.

38. The primary question is now whether the energy principle is still valid after including equations (86), (87), (88). So we will multiply equation (87) by the components of a three dimensional vector, say $\mathfrak{a}$, similarly equation (86) by a three-dimensional scalar s, and add both equations. The terms containing derivatives with respect to the coordinates are then:

$$\mathfrak{a}_x \cdot \frac{\partial u}{\partial x} + \mathfrak{a}_y \cdot \frac{\partial u}{\partial y} + \mathfrak{a}_z \cdot \frac{\partial u}{\partial z} + s \cdot \left(\frac{\partial \mathfrak{k}_x}{\partial x} + \frac{\partial \mathfrak{k}_y}{\partial y} + \frac{\partial \mathfrak{k}_z}{\partial z} \right).$$

For this expression to represent a divergence, $\mathfrak{a} = \mathfrak{k}$, $s = u$ must hold. Thus we have found for the last part of the energy equation:

$$\operatorname{div}(u\mathfrak{k}) + \mathfrak{k} \cdot \frac{\partial \mathfrak{g}}{\partial t} + u \cdot \frac{\partial w}{\partial t} - \gamma \cdot H \cdot \frac{\partial \omega}{\partial t} = 0,$$

where in the last term $\partial w / \partial t$ was substituted for u, according to equation (88).

So including the effects of gravitation results in the *total energy current* (instead of I, equation (5) on p. 641 [p. 522 in the original]):

$$\mathfrak{s} = [\mathfrak{e} \cdot \mathfrak{h}] - \varphi \cdot \mathfrak{v} + u \cdot \mathfrak{k} \tag{89}$$

and the *total change of the energy density*:

$$dW = \mathfrak{e} \cdot d\mathfrak{d} + \mathfrak{h} \cdot d\mathfrak{b} - \varphi d\rho - \mathfrak{v} \cdot d\mathfrak{f} + \mathfrak{k} \cdot d\mathfrak{g} + u dw - \gamma H d\omega. \tag{90}$$

The function H must now be defined by the following equation, instead of equation (7) of I on p. 641, [p. 523 in the original]:

$$W = H + \mathfrak{h} \cdot \mathfrak{b} - \mathfrak{v} \cdot \mathfrak{f} + uw. \tag{91}$$

[30] | It then follows from (90):

$$dH = \mathfrak{e} \cdot d\mathfrak{d} - \mathfrak{b} \cdot d\mathfrak{h} - \varphi d\rho + \mathfrak{f} \cdot d\mathfrak{v} + \mathfrak{k} \cdot d\mathfrak{g} - w du - \gamma H d\omega. \tag{92}$$

Because H is a function of the following variables: $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u, \omega)$, we have:

$$\mathfrak{e} = \frac{\partial H}{\partial \mathfrak{d}}, \quad \mathfrak{b} = -\frac{\partial H}{\partial \mathfrak{h}}, \quad \varphi = -\frac{\partial H}{\partial \rho}, \quad \mathfrak{f} = \frac{\partial H}{\partial \mathfrak{v}}, \quad \mathfrak{k} = \frac{\partial H}{\partial \mathfrak{g}}, \quad w = -\frac{\partial H}{\partial u}, \quad \frac{\partial H}{\partial \omega} = -\gamma H. \tag{93}$$

From the last equation of (93) it follows that:

$$H = e^{-\gamma\omega} H'(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u). \tag{94}$$

If we now define:

$$\mathfrak{e}' = \frac{\partial H'}{\partial \mathfrak{d}}, \quad \mathfrak{b}' = -\frac{\partial H'}{\partial \mathfrak{h}}, \quad \mathfrak{p}' = -\frac{\partial H'}{\partial \mathfrak{p}}, \quad \mathfrak{f}' = \frac{\partial H'}{\partial \mathfrak{v}}, \quad \mathfrak{k}' = \frac{\partial H'}{\partial \mathfrak{g}}, \quad \mathfrak{w}' = -\frac{\partial H'}{\partial u}, \qquad (95)$$

where all the primed quantities depend only on $(\mathfrak{d}, \mathfrak{h}, \mathfrak{p}, \mathfrak{v}, \mathfrak{g}, u)$ but not on ω, then we have:

$$\mathfrak{e} = e^{-\gamma\omega}\mathfrak{e}', \quad \mathfrak{b} = e^{-\gamma\omega}\mathfrak{b}', \quad \mathfrak{p} = e^{-\gamma\omega}\mathfrak{p}', \quad \mathfrak{f} = e^{-\gamma\omega}\mathfrak{f}', \quad \mathfrak{k} = e^{-\gamma\omega}\mathfrak{k}', \quad \mathfrak{w} = e^{-\gamma\omega}\mathfrak{w}'. \quad (96)$$

If equations (93) are satisfied, then the energy principle is valid also for the extended basic equations, and if all variables occur in H only in combinations that are invariant under Lorentz transformations, then the principle of relativity is also valid.

Thus we succeeded in devising a theory of gravitation in which both the energy principle and the principle of relativity are valid.

I want to stress particularly the last, because in the theory of matter here proposed an ansatz contradicting the principle of relativity should be rejected outright. In his papers on gravitation M. Abraham advocates the view[17] that gravitation and | relativity theory are not compatible with each other. If this were the case one would have to conclude that gravitation is so to speak a purely external force, which plays no part in the existence of matter itself. For if it belonged, as I assume here, to the forces that determine in an essential way the form of the material elementary particles and the whole internal structure of the atoms, and if it did not obey the principle of relativity, then it would be unthinkable that the elementary particles of matter and the action of forces that bind them into atoms, molecules, and tangible bodies should, when the matter moves through space, quite generally be subjected to precisely those changes that lead to the contraction of matter, which was proved by Michelson's experiment. On the other hand, however, I also believe that one would encounter great difficulties if one wanted to treat gravitation as an action that did not play any appreciable role in the internal processes of atoms, and hence I believe that one must abandon M. Abraham's point of view as soon as one treats the theory of gravitation not detached from the theory of matter. Therefore it seems to me very important that gravitation and relativity theory can be joined together in such a simple way as we have just done. [31]

Let me add the remark that in the dynamics of the aether when extended by equations (86), (87), (88), *Hamilton's principle* is valid in the form that we encountered in I, section 10. The proof offers no difficulties whatsoever.

The Invariants

39. The number of invariants is considerably increased by including the gravitational quantities. Besides the gravitational potential ω, four further quantities join the four quantities found in (24) of I. on p. 648 [p. 531 in the original], so the function H' in

17 M. Abraham, *Ann. d. Phys*. 38. p. 1056. 1912.

(94) can possibly depend on *eight* independent variables. These can be taken as the
[32] following combinations of the variables of state: |

$$\left.\begin{aligned} p &= \mathfrak{d}^2 - \mathfrak{h}^2, \\ q &= (\mathfrak{d} \cdot \mathfrak{h}), \\ \sigma &= \sqrt{\rho^2 - \mathfrak{v}^2}, \\ s &= (\rho\mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}])^2 - (\mathfrak{v} \cdot \mathfrak{h})^2, \\ \kappa &= \sqrt{\mathfrak{g}^2 - u^2}, \\ \mathfrak{k} &= (u\mathfrak{h} - [\mathfrak{g} \cdot \mathfrak{d}])^2 - (\mathfrak{g} \cdot \mathfrak{h})^2, \\ h &= (\mathfrak{g} \cdot \mathfrak{v}) - u\rho, \\ \mathfrak{b} &= (\rho\mathfrak{h} - [\mathfrak{v} \cdot \mathfrak{d}]) \cdot (u\mathfrak{h} - [\mathfrak{g} \cdot \mathfrak{d}]) - (\mathfrak{v} \cdot \mathfrak{h}) \cdot (\mathfrak{g} \cdot \mathfrak{h}). \end{aligned}\right\} \tag{97}$$

It can be proved by means of four-dimensional vector analysis that all other invariants can be computed from these eight quantities. But I do not want to reproduce that proof here.

Similarly I do not wish to write down here the formulas that now lead to the calculation of the quantities $\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', w', \mathfrak{k}'$ from the function H', analogous to the formulae (25) in I, p. 649 [p. 531 in the original], since they can be derived quite easily.

The Differential Equation of the Electron

40. The following quantities are of course also invariants under Lorentz transformations:

$$\begin{aligned} \mathfrak{e} \cdot \mathfrak{d} - \mathfrak{b} \cdot \mathfrak{h} &= e^{-\gamma\omega}(\mathfrak{e}' \cdot \mathfrak{d} - \mathfrak{b}' \cdot \mathfrak{h}), \\ \varphi\rho - \mathfrak{f} \cdot \mathfrak{v} &= e^{-\gamma\omega}(\varphi'\rho - \mathfrak{f}' \cdot \mathfrak{v}), \\ \mathfrak{k} \cdot \mathfrak{g} - wu &= e^{-\gamma\omega}(\mathfrak{k}' \cdot \mathfrak{g} - w'u). \end{aligned}$$

For many purposes it is more convenient to use other functions instead of H, which differ from the latter only by an additional term formed from the quantities just written down. Let us define exactly as in I, p. 642 [p. 524 in the original]:

$$\Phi = H - (\mathfrak{e} \cdot \mathfrak{d} - \mathfrak{b} \cdot \mathfrak{h}) + (\varphi\rho - \mathfrak{f} \cdot \mathfrak{v}). \tag{98}$$

We can also set:

$$\left.\begin{aligned} \Phi &= e^{-\gamma\omega}\Phi', \\ \Phi' &= H' - (\mathfrak{e}' \cdot \mathfrak{d} - \mathfrak{b}' \cdot \mathfrak{h}) + (\varphi'\rho - \mathfrak{f}' \cdot \mathfrak{v}), \end{aligned}\right\} \tag{99}$$

where accordingly Φ' is a quantity depending only on the variables $\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u$
[33] and not on ω. Since $(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}')$ | can be calculated from the variables

$(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u)$, one can also, conversely, calculate $(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v})$ from $(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', \mathfrak{g}, u)$, and so one may consider Φ' as a function of this new system of variables:

$$\Phi = e^{-\gamma\omega}\Phi'(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', \mathfrak{g}, u). \tag{100}$$

Now it follows from (99) and (95):

$$d\Phi' = -\mathfrak{d} \cdot d\mathfrak{e}' + \mathfrak{h} \cdot d\mathfrak{b}' + \rho d\varphi' - \mathfrak{v} \cdot d\mathfrak{f}' + \mathfrak{k}' \cdot d\mathfrak{g} - w' du,$$

therefore:

$$\mathfrak{d} = -\frac{\partial\Phi'}{\partial\mathfrak{e}'}, \quad \mathfrak{h} = \frac{\partial\Phi'}{\partial\mathfrak{b}'}, \quad \rho = \frac{\partial\Phi'}{\partial\varphi'}, \quad \mathfrak{v} = -\frac{\partial\Phi'}{\partial\mathfrak{f}'}, \quad \mathfrak{k}' = \frac{\partial\Phi'}{\partial\mathfrak{g}}, \quad w' = -\frac{\partial\Phi'}{\partial u}. \tag{101}$$

In the case of an electron at rest the quantities $\mathfrak{b}', \mathfrak{f}', u$ are to be set to constant zero, and the three remaining ones depend only on the distance r from the center. I set:

$$\left.\begin{aligned} X &= \mathfrak{e}' = -e^{+\gamma\omega}\frac{d\varphi}{dr}, \\ Y &= \varphi' = e^{+\gamma\omega}\varphi, \\ Z &= \mathfrak{g} = \frac{d\omega}{dr}. \end{aligned}\right\} \tag{102}$$

Thus we have a function Φ' that depends only on three variables:

$$\Phi'(X, Y, Z).$$

But because also:

$$X = -\frac{dY}{dr} + \gamma YZ,$$

we really have in Φ' only two unknown variables Y and Z and the derivative dY/dr of one of them.

For these two unknown variables Y and Z we then have the following two differential equations:

$$\frac{1}{r^2}\frac{d}{dr}(r^2\mathfrak{d}) = \rho,$$

$$\frac{1}{r^2}\frac{d}{dr}(r^2\mathfrak{k}) = -\gamma H,$$

or:

$$\left.\begin{aligned} &\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{\partial\Phi'}{\partial X}\right)+\frac{\partial\Phi'}{\partial Y}\;=\;0,\\ &\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{\partial\Phi'}{\partial Z}\right)+\gamma\cdot\left(\Phi'-X\,\frac{\partial\Phi'}{\partial X}-Y\,\frac{\partial\Phi'}{\partial Y}-Z\,\frac{\partial\Phi'}{\partial Z}\right)\;=\;0. \end{aligned}\right\} \tag{103}$$

[34] | These two equations should replace equation (34) of II, p. 15 when one wants to discuss the problem of the electron with gravitation taken into account. Incidentally, the unknown Z and its derivatives can also be eliminated from both equations according to the usual procedure of differential calculus. This yields an equation of third order for the unknown $Y = e^{\gamma\omega}\varphi$, whereas (34) was an equation of second order for φ.

The World Matrix

41. As before (I, p. 643 [p. 525 in the original], we can use the world function Φ defined in the previous section to construct the world matrix. Namely, from equation (98) and (91):

$$W\;=\;\Phi+\mathfrak{e}\cdot\mathfrak{d}-\varphi\rho+uw. \tag{104}$$

The world matrix can now be constructed according to exactly the same scheme as in equation (16), simply by including the four-vector of gravitation:

$$\left|\begin{array}{c} \Phi-\mathfrak{b}\cdot\mathfrak{h}+\mathfrak{e}_x\mathfrak{d}_x+\mathfrak{h}_x\mathfrak{b}_x+\mathfrak{f}_x\mathfrak{v}_x-\mathfrak{g}_x\mathfrak{k}_x,\\ \mathfrak{e}_x\mathfrak{d}_y+\mathfrak{h}_x\mathfrak{b}_y+\mathfrak{f}_x\mathfrak{v}_y-\mathfrak{g}_x\mathfrak{k}_y,\quad \mathfrak{e}_x\mathfrak{d}_z+\mathfrak{h}_x\mathfrak{b}_z+\mathfrak{f}_x\mathfrak{v}_z-\mathfrak{g}_x\mathfrak{k}_z,\\ -i(\mathfrak{d}_y\mathfrak{b}_z-\mathfrak{d}_z\cdot\mathfrak{b}_y-\mathfrak{f}_x\rho+\mathfrak{g}_x w),\\ \mathfrak{e}_y\mathfrak{d}_x+\mathfrak{h}_y\mathfrak{b}_x+\mathfrak{f}_y\mathfrak{v}_x-\mathfrak{g}_y\mathfrak{k}_x,\\ \Phi-\mathfrak{b}\cdot\mathfrak{h}+\mathfrak{e}_y\mathfrak{d}_y+\mathfrak{h}_y\mathfrak{b}_y+\mathfrak{f}_y\mathfrak{v}_y-\mathfrak{g}_y\mathfrak{k}_y,\\ \mathfrak{e}_y\mathfrak{d}_z+\mathfrak{h}_y\mathfrak{b}_z+\mathfrak{f}_y\mathfrak{v}_z-\mathfrak{g}_y\mathfrak{k}_z,\quad -i(\mathfrak{d}_z\mathfrak{b}_x-\mathfrak{d}_x\mathfrak{b}_z-\mathfrak{f}_y\rho+\mathfrak{g}_y w),\\ \mathfrak{e}_z\mathfrak{d}_x+\mathfrak{h}_z\mathfrak{b}_x+\mathfrak{f}_z\mathfrak{v}_x-\mathfrak{g}_z\mathfrak{k}_x,\quad \mathfrak{e}_z\mathfrak{d}_y+\mathfrak{h}_z\mathfrak{b}_y+\mathfrak{f}_z\mathfrak{v}_y-\mathfrak{g}_z\mathfrak{k}_y,\\ \Phi-\mathfrak{b}\cdot\mathfrak{h}+\mathfrak{e}_z\mathfrak{d}_z+\mathfrak{h}_z\mathfrak{b}_z+\mathfrak{f}_z\mathfrak{v}_z-\mathfrak{g}_z\mathfrak{k}_z,\\ -i(\mathfrak{d}_x\mathfrak{b}_y-\mathfrak{d}_y\mathfrak{b}_x-\mathfrak{f}_z\rho+\mathfrak{g}_z w),\\ -i(\mathfrak{e}_y\mathfrak{h}_z-\mathfrak{e}_z\mathfrak{h}_y-\mathfrak{v}_x\varphi+\mathfrak{k}_x u),\\ -i(\mathfrak{e}_z\mathfrak{h}_x-\mathfrak{e}_y\mathfrak{h}_z-\mathfrak{v}_y\varphi+\mathfrak{k}_y u),\\ -i(\mathfrak{e}_x\mathfrak{h}_y-\mathfrak{e}_y\mathfrak{h}_x-\mathfrak{v}_z\varphi+\mathfrak{k}_z u),\quad \Phi+\mathfrak{e}\cdot\mathfrak{d}-\varphi\rho+uw. \end{array}\right| \tag{105}$$

When the operation Δtv is applied to this matrix, the fourth row yields the energy principle; the first three rows will lead to the equations of motion of a material particle. |

For the most general case that all variables (97) occur in H, and by quite elementary calculational steps, of hardly greater complexity than those mentioned in section 25, one can prove the equations: [35]

$$[\mathfrak{e} \cdot \mathfrak{d}] + [\mathfrak{h} \cdot \mathfrak{b}] + [\mathfrak{f} \cdot \mathfrak{v}] + [\mathfrak{k} \cdot \mathfrak{g}] = 0, \tag{106}$$

$$[\mathfrak{e} \cdot \mathfrak{h}] + [\mathfrak{b} \cdot \mathfrak{d}] + (\rho \mathfrak{f} - \varphi \mathfrak{v}) + (u \mathfrak{k} - w \mathfrak{g}) = 0, \tag{107}$$

which are the generalizations of equations (54) and (55), so that

$$\mathfrak{e}_x \mathfrak{d}_y + \mathfrak{h}_x \mathfrak{b}_y + \mathfrak{f}_x \mathfrak{v}_y - \mathfrak{g}_x \mathfrak{k}_y = \mathfrak{d}_x \mathfrak{e}_y + \mathfrak{b}_x \mathfrak{h}_y + \mathfrak{v}_x \mathfrak{f}_y - \mathfrak{k}_x \mathfrak{g}_y \quad \text{etc.}$$
$$\mathfrak{d}_y \mathfrak{b}_z - \mathfrak{d}_z \mathfrak{b}_y - \rho \mathfrak{f}_x + w \mathfrak{g}_x = \mathfrak{e}_y \mathfrak{h}_z - \mathfrak{e}_z \mathfrak{h}_y - \varphi \mathfrak{v}_x + u \mathfrak{k}_x \quad \text{etc.}$$

The world matrix (105) is symmetric across the diagonal.

Calculation of the Force Acting on a Mass Particle

42. To calculate the ponderomotive force on a mass particle we proceed exactly as in 26. We consider a volume V containing the mass particle that on the one hand is large enough that the law of superposition, valid in vacuo, is already satisfied on its surface S; but which is on the other hand small enough that the largely extended field, which causes the force action, can be well approximated as homogeneous in the interior, if the particle is imagined to be absent.

For the actions of gravity, the principle of superposition states that the differential equations (86) and (87) are linear in $\mathfrak{g}$ and u. Therefore this condition must be fulfilled:

In vacuo $(\mathfrak{g}, iu)$ *and* $(\mathfrak{k}, iw)$ *differ only by a constant factor.*

The factor of proportionality depends only on how we define the units.[18] We will make the convention that *in vacuo*:

$$\mathfrak{g} = \mathfrak{k}. \tag{108}$$

| Accordingly the world function Φ must be representable on the surface S of the volume V in the form: [36]

$$\Phi = \frac{1}{2}(\mathfrak{b}^2 - \mathfrak{e}^2) + \frac{1}{2}(\mathfrak{g}^2 - u^2) + \Phi_1, \tag{109}$$

where Φ_1 can be neglected as vanishingly small, and with an error that decreases as V is taken to be increasingly large.

Let $\mathfrak{s}$ be the energy current, then the value of the momentum of motion $\mathfrak{G}$ contained in V is given by:

---

18 So that we can disregard the factor $e^{-\gamma\omega}$, we assume that the gravitational potential ω is so small that $e^{-\gamma\omega}$ cannot be distinguished from 1. We will see in 47. that this assumption does not restrict the general validity of the proofs.

$$\mathfrak{G} = \int_V \mathfrak{s} dV.$$

The inertial mass M contained in the volume V amounts to:

$$M = \int_V W dV,$$

where W denotes the density of energy. Therefore the velocity $\mathfrak{q}$ of the considered particle is:

$$\mathfrak{q} = \frac{\mathfrak{G}}{M}.$$

Then the force $\mathfrak{P}$ acting on the particle is calculated to be:

$$\mathfrak{P} = \frac{d\mathfrak{G}}{dt} = \int_V \frac{\partial \mathfrak{s}}{\partial t} dV + \int_V (\mathfrak{q} \cdot \nabla)\mathfrak{s} dV.$$

Here the first term on the right side signifies the change in time of the momentum of motion in the stationary volume V, in which the particle moves so that $\mathfrak{s}$ changes; and the second term means the change in momentum calculated for a volume V in which the particle remains at rest and $\mathfrak{s}$ is unchanged, but where V is displaced with velocity $\mathfrak{q}$. The two together yield the change in time of $\mathfrak{G}$ in a co-moving volume V rigidly attached to the particle.

We now substitute for $\partial \mathfrak{s}/\partial t$ the values calculated from equation (58) and obtain:

$$\mathfrak{P} = \int_S \mathfrak{p}_N dS + \int_S \mathfrak{s} q dS \cdot (N, \mathfrak{q}).$$

by a single integration (cf. equation (62)). |

[37] Since the principle of superposition is valid on S, the components of $\mathfrak{p}_N$ as well as those of $\mathfrak{s}$ are expressions of second degree in the state variables; that is, $\mathfrak{P}$ is composed additively of an expression that contains only the electromagnetic field quantities (the force of the electromagnetic field), and of an expression that contains only the gravitational quantities $\mathfrak{g}$ and u (the gravitational force acting on the particle). Since we already know the first expression we are interested here only in the second. So we calculate with the expressions:

$$\mathfrak{p}_{1x} = -\frac{1}{2}(\mathfrak{g}_x^2 - \mathfrak{g}_y^2 - \mathfrak{g}_z^2) - \frac{1}{2}u^2,$$
$$\mathfrak{p}_{1y} = -\mathfrak{g}_x\mathfrak{g}_y,$$
$$\mathfrak{p}_{1z} = -\mathfrak{g}_x\mathfrak{g}_z,$$
$$\mathfrak{s}_x = \mathfrak{g}_x u,$$

which result from the matrix (105) if one sets $\mathfrak{k} = \mathfrak{g}$, $u = w$ and the value (109) for Φ, and furthermore omits all terms in $\mathfrak{b}$ and $\mathfrak{e}$.

Further, according to the principle of superposition we can now think of the several components of the variables of state composed additively of two quantities each; one which corresponds to the particle's proper field, to be denoted by the index 0, and a quantity belonging to the largely extended field in which the particle moves, denoted by the index 1:

$$\mathfrak{g} = \mathfrak{g}_0 + \mathfrak{g}_1, \quad u = u_0 + u_1.$$

For the tensor components $\mathfrak{p}$ and the components of $\mathfrak{s}$ we obtain sums of three expressions each, the first of which is composed only of quantities with index 0, the second of quantities with mixed indices, and the third of quantities with index 1:

$$\mathfrak{p}_{1x} = -\frac{1}{2}(\mathfrak{g}_{x0}^2 - \mathfrak{g}_{y0}^2 - \mathfrak{g}_{z0}^2 + u_0^2) - (\mathfrak{g}_{x0}\mathfrak{g}_{x1} - \mathfrak{g}_{y0}\mathfrak{g}_{y1} - \mathfrak{g}_{z0}\mathfrak{g}_{z1} + u_0u_1)$$
$$-\frac{1}{2}(\mathfrak{g}_{x1}^2 - \mathfrak{g}_{y1}^2 - \mathfrak{g}_{z1}^2 + u_1^2),$$
$$\mathfrak{p}_{1y} = -\mathfrak{g}_{x0}\mathfrak{g}_{y0} - (\mathfrak{g}_{x1}\mathfrak{g}_{y0} + \mathfrak{g}_{x0}\mathfrak{g}_{y1}) - \mathfrak{g}_{x1}\mathfrak{g}_{y1},$$
$$\mathfrak{p}_{1z} = -\mathfrak{g}_{x0}\mathfrak{g}_{z0} - (\mathfrak{g}_{x1}\mathfrak{g}_{z0} + \mathfrak{g}_{x0}\mathfrak{g}_{z1}) - \mathfrak{g}_{x1}\mathfrak{g}_{z1},$$
$$\mathfrak{s}_x = u_0\mathfrak{g}_{x0} + (u_1\mathfrak{g}_{x0} + u_0\mathfrak{g}_{x1}) + u_1\mathfrak{g}_{x1}.$$

| Correspondingly, $\mathfrak{P}$ decomposes into three terms as well, which one could [38]
denote by $\mathfrak{P}_{00}, \mathfrak{P}_{01}, \mathfrak{P}_{11}$. $\mathfrak{P}_{00}$ would be obtained by annulling the extended field in which the particle is moving ($\mathfrak{g}_1 = 0$, $u_1 = 0$). But because the particle's proper field is in internal equilibrium with itself, the particle can move only with constant velocity in a field-free space ($\mathfrak{g}_1 = 0$, $u_1 = 0$), therefore $\mathfrak{P}_{00} = 0$. In exactly the same way we find $\mathfrak{P}_{11} = 0$. So to calculate $\mathfrak{P}$ there remain only the terms that we have characterized as $\mathfrak{P}_{01}$. For example, the result for the x-component of the force is:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int_S(\mathfrak{g}_{x0}dS_x + \mathfrak{g}_{y0}dS_y + \mathfrak{g}_{z0}dS_z) - u_1\int_S u_0 dS_x$$

$$+\mathfrak{g}_{y1}\int_S(\mathfrak{g}_{y0}dS_x - \mathfrak{g}_{x0}dS_y) + \mathfrak{g}_{z1}\int_S(\mathfrak{g}_{z0}dS_x - \mathfrak{g}_{x0}dS_z)$$

$$+ u_1\int_S \mathfrak{g}_{x0}(\mathfrak{q}_x dS_x + \mathfrak{q}_y dS_y + \mathfrak{q}_z dS_z) + \mathfrak{g}_{x1}\int_S u_0(\mathfrak{q}_x dS_x + \mathfrak{q}_y dS_y + \mathfrak{q}_z dS_z).$$

Here the assumption was used that the extended field $\mathfrak{g}_1, u_1$ may be treated as constant in the interior of the volume V. Further we have set for brevity:

$$dS\cos(N,x) = dS_x, \qquad dS\cos(N,y) = dS_y, \qquad dS\cos(N,z) = dS_z.$$

From the property of the vector $\mathfrak{g}$ that (from equation (87)):

$$\operatorname{curl}\mathfrak{g} = 0,$$

it follows:

$$\int_S(\mathfrak{g}_{y0}dS_x - \mathfrak{g}_{x0}dS_y) = 0,$$

$$\int_S(\mathfrak{g}_{z0}dS_x - \mathfrak{g}_{x0}dS_z) = 0.$$

[39] So this eliminates the third and fourth term in the sum for $\mathfrak{P}_x$ written above. I again change the remaining terms into volume integrals, making simultaneously multiple use of $\mathfrak{k}_0 = \mathfrak{g}_0$, $w_0 = u_0$ which are valid on S:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int\left(\frac{\partial \mathfrak{k}_{x0}}{\partial x} + \frac{\partial \mathfrak{k}_{y0}}{\partial y} + \frac{\partial \mathfrak{k}_{z0}}{\partial z}\right)dV - u_1\int\frac{\partial u_0}{\partial x}dV$$

$$+ u_1\int\left(\frac{\partial \mathfrak{g}_{x0}}{\partial x}\mathfrak{q}_x + \frac{\partial \mathfrak{g}_{x0}}{\partial y}\mathfrak{q}_y + \frac{\partial \mathfrak{g}_{x0}}{\partial z}\mathfrak{q}_z\right)dV + \mathfrak{g}_{x1}\int\left(\frac{\partial w_0}{\partial x}\mathfrak{q}_x + \frac{\partial w_0}{\partial y}\mathfrak{q}_y + \frac{\partial w_0}{\partial z}\mathfrak{q}_z\right)dV.$$

But we have:

$$\frac{\partial \mathfrak{g}_{x0}}{\partial x}\mathfrak{q}_x + \frac{\partial \mathfrak{g}_{x0}}{\partial y}\mathfrak{q}_y + \frac{\partial \mathfrak{g}_{x0}}{\partial z}\mathfrak{q}_z = \frac{d\mathfrak{g}_{x0}}{dt} - \frac{\partial \mathfrak{g}_{x0}}{\partial t} = \frac{d\mathfrak{g}_{x0}}{dt} + \frac{\partial u_0}{\partial x}$$

and:

$$\frac{\partial w_0}{\partial x}\mathfrak{q}_x + \frac{\partial w_0}{\partial y}\mathfrak{q}_y + \frac{\partial w_0}{\partial z}\mathfrak{q}_z = \frac{dw_0}{dt} - \frac{\partial w_0}{\partial t}.$$

so that the result is:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int\left(\frac{\partial \mathfrak{k}_{x0}}{\partial x} + \frac{\partial \mathfrak{k}_{y0}}{\partial y} + \frac{\partial \mathfrak{k}_{z0}}{\partial z} + \frac{\partial w_0}{\partial t}\right)dV + \int\left(w_1\frac{d\mathfrak{g}_{x0}}{dt} + \mathfrak{g}_{x1}\frac{dw_0}{dt}\right)dV.$$

The second term is vanishingly small compared to the first. For since w_1 and $\mathfrak{g}_{x1}$ are vanishingly small compared to w_0 and $\mathfrak{g}_{x0}$ in the interior of the volume V, that second term is negligible compared to the term:

$$\frac{d}{dt}\int w\mathfrak{g}_x dV,$$

which occurs in another formula for $\mathfrak{P}_x$, namely:

$$\mathfrak{P}_x = \frac{d\mathfrak{G}_x}{dt} = \frac{d}{dt}\int(\mathfrak{d}_y\mathfrak{b}_z - \mathfrak{d}_z\mathfrak{b}_y - \mathfrak{f}_x\rho + \mathfrak{g}_x w)dV$$

and so that term must be negligible compared to the value of $\mathfrak{P}_x$ in general. Thus we have:

$$\mathfrak{P}_x = -\mathfrak{g}_{x1}\int_V(\frac{\partial \mathfrak{k}_{x0}}{\partial x} + \frac{\partial \mathfrak{k}_{y0}}{\partial y} + \frac{\partial \mathfrak{k}_{z0}}{\partial z} + \frac{\partial w_0}{\partial t})dV.$$

In the following I again omit the indices and put according to (86):

$$\frac{\partial \mathfrak{k}_{x0}}{\partial x} + \frac{\partial \mathfrak{k}_{y0}}{\partial y} + \frac{\partial \mathfrak{k}_{z0}}{\partial z} + \frac{\partial w_0}{\partial t} = -\gamma H,$$

| where H now denotes the Hamiltonian function of the particle's proper field. I also [40]
put $\mathfrak{g}_1 = \mathfrak{g}$, that is by $\mathfrak{g}$ I mean the field strength of the extended field in which the particle is moving. Then:

$$\mathfrak{P} = \gamma\mathfrak{g}\int_V H dV. \tag{110}$$

We have to define the *gravitational mass m_g of the particle* as:

$$m_g = \int_V H dV \tag{111}$$

and then:

$$\mathfrak{P} = \gamma m_g \mathfrak{g}. \tag{112}$$

We can therefore state the following proposition:

In a gravitational field there is only one type of action of force, the action of gravity, and there is nothing that would be related to it as the magnetic action of force is related to the electric one. But the gravitational mass of a material particle depends on its state of motion, in contrast to the constancy of the electric charge.

For, we have seen on p. 663 [p. 26 in the original] that:

$$\int H dV = \sqrt{1-q^2}E_0,$$

$$m_g = \sqrt{1-q^2}m_0, \tag{113}$$

if we understand m_0 to be the gravitational mass of the particle at rest. Moreover it is straightforward to verify the validity of the following proposition:

For a massive particle at rest the gravitational mass and the inertial mass are identical.

Both are $m_0 = E_0$. To what extent they differ in a moving body will be seen in a later section (45.).

Now I consider two particles at rest or in slow motion, having masses m_{g1} and m_{g2}. In the surrounding empty space the field strength of gravitation is calculated to be, due to $\operatorname{div}\mathfrak{k} = -\gamma H$ and, in vacuo, $\mathfrak{g} = \mathfrak{k}$:

$$\mathfrak{g}_1 = \gamma\frac{m_{g1}}{4\pi {r_1}^2}, \quad \mathfrak{g}_2 = \gamma\frac{m_{g2}}{4\pi {r_2}^2},$$

[41] | where r_1 and r_2 shall denote the radius vectors from the particles, and where the field lines point toward the particle generating the field. If the two are at a distance $r = r_1 = r_2$ from each other they accordingly attract each other with a force of equal magnitude:

$$\mathfrak{P}_g = \frac{\gamma^2}{4\pi}\frac{m_{g1}m_{g2}}{r^2}. \tag{114}$$

In our theory of gravitation the law of equality of action and reaction as well as Newton's law of attraction are valid.

Both laws are a necessary consequence of the principle of superposition that holds in vacuo.

Because the superposition principle also implies that the gravitational fields of very many mass particles, which merely represent the sinks of the vector $\mathfrak{k}$, simply add together, it follows *that the attractive effect of a body in the universe is altered in no way by interposing another body; rather the effect of the second body superposes unchanged; in other words, gravity cannot be shielded*.

The energy density, calculated according to the formula:

$$W = \Phi + \mathfrak{e}\cdot\mathfrak{d} - \varphi\rho + u w$$

becomes the following in vacuo, where $\mathfrak{e} = \mathfrak{d}, \mathfrak{h} = \mathfrak{b}, \rho = 0, \mathfrak{v} = 0, \mathfrak{k} = \mathfrak{g}, w = u$:

$$W = \frac{1}{2}(\mathfrak{e}^2 + \mathfrak{h}^2 + \mathfrak{g}^2 + u^2).$$

The energy density of the gravitational field in vacuo is a positive quantity.

This theorem, which is remarkable considering the attractive nature of the gravitational field, has already been derived by M. Abraham from his ansatz for the equations of gravitation.

Finally we want to calculate the numerical value of the universal constant γ from the above Newtonian law of attraction (114). In formula (114) the two gravitational masses m_{g1} and m_{g2} are to be expressed in ergs. | First let us give them in the usual [42] fashion in grams by putting:

$$m_1 = \frac{m_{g1}}{c^2}, \quad m_2 = \frac{m_{g2}}{c^2},$$

where c is the speed of light $(3 \cdot 10^{10})$. The law of attraction then takes the form:

$$\mathfrak{P}_g = \frac{\gamma^2 c^4}{4\pi} \frac{m_1 m_2}{r^2}. \tag{114a}$$

We denote what is usually called the *gravitational constant* by κ, so we have:

$$\kappa = \frac{\gamma^2 c^4}{4\pi}. \tag{115}$$

Therefore:

$$\gamma = \frac{\sqrt{4\pi\kappa}}{c^2}.$$

When we substitute:

$$\kappa = 6,648 \cdot 10^{-8},$$

we obtain:

$$\gamma = 1,016 \cdot 10^{-24}.$$

The Inertial Mass of a Material Particle

43. If E_0 is the energy of a particle at rest, then the energy of the same particle in motion with speed q follows from Laue's theorem:[19]

$$E = \frac{E_0}{\sqrt{1 - q^2}}.$$

According to (91) and (98) the energy density is:

$$\begin{aligned} W &= H + \mathfrak{b} \cdot \mathfrak{h} - \mathfrak{f} \cdot \mathfrak{v} + wu \\ &= \Phi + \mathfrak{e} \cdot \mathfrak{d} - \varphi\rho + wu, \end{aligned}$$

19 M. Laue, *Das Relativitätsprinzip*, p. 170.

hence, considering (64) and (65):

$$E_0 = \int(\overline{H_0} + \overline{w_0 u_0})dV_0 = \int(\overline{\Phi_0} + \overline{w_0 u_0})dV_0. \tag{116}$$

These formulas now take the place of the formulas (66). |

[43] The following formula can be easily derived from equations (85), (86):

$$\mathfrak{k} \cdot \mathfrak{g} - wu = \operatorname{div}(\mathfrak{k}\omega) + \frac{\partial(w\omega)}{\partial t} + \gamma\omega H.$$

This implies for a material particle *at rest*, by integration over the whole volume V_0 it occupies:

$$\int_{V_0} \overline{\mathfrak{k}_0 \cdot \mathfrak{g}_0} dV_0 - \int_{V_0} \overline{w_0 u_0} dV_0 = \gamma \int_{V_0} \overline{\omega_0 H_0} dV_0. \tag{117}$$

In place of the three equations on p. 659 [p. 10 in the original] we obtain the following equations from Laue's theorem, taking into account the gravitational terms:

$$\begin{aligned}
\int(\overline{\mathfrak{v}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{x0}\mathfrak{h}_{x0}} - \overline{\mathfrak{e}_{x0}\mathfrak{d}_{x0}} - \overline{\mathfrak{f}_{x0}\mathfrak{v}_{x0}} + \overline{\mathfrak{k}_{x0}\mathfrak{g}_{x0}} + \overline{w_0 u_0})dV_0 &= E_0, \\
\int(\overline{\mathfrak{v}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{y0}\mathfrak{h}_{y0}} - \overline{\mathfrak{e}_{y0}\mathfrak{d}_{y0}} - \overline{\mathfrak{f}_{y0}\mathfrak{v}_{y0}} + \overline{\mathfrak{k}_{y0}\mathfrak{g}_{y0}} + \overline{w_0 u_0})dV_0 &= E_0, \\
\int(\overline{\mathfrak{v}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{b}_{z0}\mathfrak{h}_{z0}} - \overline{\mathfrak{e}_{z0}\mathfrak{d}_{z0}} - \overline{\mathfrak{f}_{z0}\mathfrak{v}_{z0}} + \overline{\mathfrak{k}_{z0}\mathfrak{g}_{z0}} + \overline{w_0 u_0})dV_0 &= E_0.
\end{aligned}$$

Addition, taking note of (65), results in:

$$E_0 = -\frac{1}{3}\int(\overline{\mathfrak{e}_0 \cdot \mathfrak{d}_0} - \overline{\mathfrak{b}_0 \cdot \mathfrak{h}_0} - \overline{\mathfrak{k}_0 \cdot \mathfrak{g}_0})dV_0 + \int\overline{w_0 u_0}dV_0 \tag{118}$$

in place of (71).

In the field of an electron we have $\mathfrak{h}_0 = 0$, $u_0 = 0$ *consequently:*

$$E_0 = -\frac{1}{3}\int\mathfrak{e}_0 \cdot \mathfrak{d}_0 dV_0 + \frac{1}{3}\int\mathfrak{k}_0 \cdot \mathfrak{g}_0 dV_0, \tag{119}$$

and here we have, from (64) and (115):

$$\int\mathfrak{e}_0 \cdot \mathfrak{d}_0 dV_0 = \int\varphi_0 \rho dV_0, \tag{120}$$

$$\int\mathfrak{k}_0 \cdot \mathfrak{g}_0 dV_0 = \gamma\int\omega_0 H_0 dV_0. \tag{121}$$

By formula (119) the proofs that we gave on p. 662 [p. 13 in the original] can no longer be carried through rigorously. Nevertheless it should be rather likely that the peculiar statements about the signs of $\mathfrak{e}$ and φ in the interior of the electron can in fact be maintained. |

Gravity of Moving Massive Particles [44]

44. Let $\mathfrak{g}_x, \mathfrak{g}_y, \mathfrak{g}_z, u$ be the gravitational field of a material particle, whose gravitational mass shall be m_0 when it is at rest. By (85), $\mathfrak{g}$ can always be derived from a gravitational potential ω:

$$\mathfrak{g}_x = \frac{\partial \omega}{\partial x}, \quad \mathfrak{g}_y = \frac{\partial \omega}{\partial y}, \quad \mathfrak{g}_z = \frac{\partial \omega}{\partial z}, \quad u = -\frac{\partial \omega}{\partial t} \quad [1]$$

and here ω is a four-dimensional scalar, that is, an invariant under Lorentz transformations. Let the particle move with speed q in the direction of the positive z-axis. We want to transform to rest, that is, we want to associate with the point x, y, z at time t a point x_0, y_0, z_0 according to the following equations:

$$x_0 = x, \quad y_0 = y, \quad z_0 = \frac{z - qt}{\sqrt{1 - q^2}}, \quad t_0 = \frac{t - qz}{\sqrt{1 - q^2}}.$$

Let the center of the material particle be the point associated with $x_0 = 0, y_0 = 0, z_0 = 0$ which therefore has the coordinates $x = 0, y = 0, z = qt$. The distance of the point x, y, z from the center of the mass particle is:

$$r = \sqrt{x^2 + y^2 + (z - qt)^2},$$

and the distance of the associated point x_0, y_0, z_0 from the center $(0, 0, 0)$ of the particle in its rest frame is:

$$r_0 = \sqrt{{x_0}^2 + {y_0}^2 + {z_0}^2}.$$

We can also calculate this quantity as a function of x, y, z, t:

$$r_0 = \sqrt{x^2 + y^2 + \frac{(z - qt)^2}{1 - q^2}}. \tag{122}$$

If we denote by ϑ the angle between the positive z-axis and the radius vector r, then:

$$z - qt = r \cos \vartheta$$

and thus we have:

$$r_0 = r \frac{\sqrt{1 - q^2 \sin^2 \vartheta}}{\sqrt{1 - q^2}} = r \sqrt{1 + \frac{q^2}{1 - q^2} \cos^2 \vartheta}.$$

[45] | Let us introduce the following abbreviation:

$$\left.\begin{aligned} &\sqrt{1+\frac{q^2}{1-q^2}\cos^2\vartheta} = p, \\ &r_0 = rp. \end{aligned}\right\} \tag{123}$$

Now the gravitational potential ω for a particle at rest is easily calculated, namely:

$$\omega = \frac{\gamma m_0}{4\pi r_0}.$$

Since ω is invariant under Lorentz transformations it follows using (122) and (123) that the potential for a moving particle is given at the point (x, y, z) by the formula:

$$\omega = \frac{\gamma m_0}{4\pi\sqrt{x^2+y^2+\frac{(z-qt)^2}{1-q^2}}} = \frac{\gamma m_0}{4\pi rp}. \tag{124}$$

This gives immediately the gravitational field of the moving particle:

$$\left.\begin{aligned} \mathfrak{g}_x &= -\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{x}{r}, \\ \mathfrak{g}_y &= -\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{y}{r}, \\ \mathfrak{g}_z &= -\frac{1}{1-q^2}\cdot\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{z-qt}{r}, \\ u &= -\frac{q}{1-q^2}\cdot\frac{\gamma}{4\pi p^3}\cdot\frac{m_0}{r^2}\cdot\frac{z-qt}{r}, \end{aligned}\right\} \tag{125}$$

or, when we denote by $\mathfrak{g}_\rho$ the component of the field normal to the direction of motion (z):

$$\left.\begin{aligned} \mathfrak{g}_\rho &= -\frac{\gamma m_0}{4\pi r^2}\cdot\frac{\sin\vartheta}{p^3}, \\ \mathfrak{g}_z &= -\frac{\gamma m_0}{4\pi r^2}\cdot\frac{\cos\vartheta}{(1-q^2)p^3}, \\ u &= -\frac{\gamma m_0 q}{4\pi r^2}\cdot\frac{\cos\vartheta}{(1-q^2)p^3}, \end{aligned}\right\} \tag{126}$$

where the value of p is to be substituted from formula (123).

The formulas (126) clearly imply the following:

The gravitational field lines in the vicinity of a material body, which extend from it in a straight | *radial direction when it is at rest, acquire a curved form when the body is in motion; in addition the field acquires sources and sinks in the vicinity of the moving particle*. [46]

The last part is easily seen: in empty space we have $\mathfrak{g} = \mathfrak{k}$ and $u = w$; since here u and also $\partial u / \partial t$ differ from zero, this is also true of $\operatorname{div}\mathfrak{g}$, for:

$$\operatorname{div}\mathfrak{g} = \operatorname{div}\mathfrak{k} = -\frac{\partial w}{\partial t} = -\frac{\partial u}{\partial t}.$$

The order of magnitude of $\operatorname{div}\mathfrak{g}$ is that of q^2, as one can easily check.

The strange distortion of the gravitational field becomes noticeable only when q takes on quite significant values. The equation for a line of force is:

$$dz : d\rho = \frac{z - qt}{1 - q^2} : \rho,$$

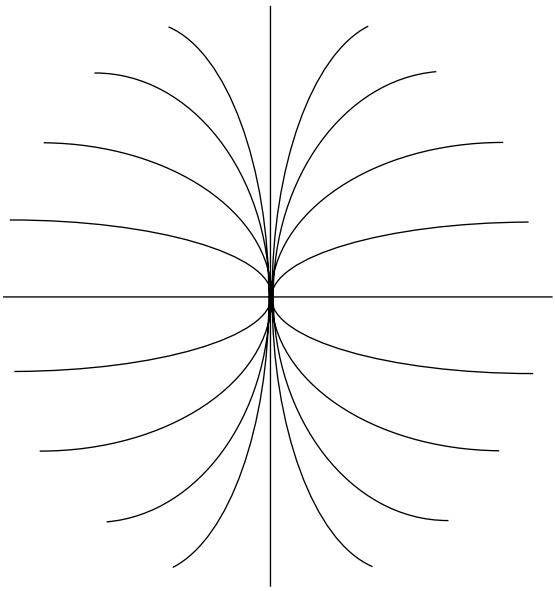

Figure 1: Shape of the gravitational field lines of a rapidly moving particle

$$\left(q = \sqrt{\frac{1}{2}} = 212000 \text{ km/sec}\right).$$

| therefore: [47]

$$z - qt = a \cdot \rho^{\frac{1}{1-q^2}},$$

where a is a parameter labeling the line of force. One sees immediately that for lesser values of q the line of force hardly deviates from the radius vector, then a is given by $\cot g\vartheta$. But if $q^2 = \frac{1}{2}$, for example, then the lines of force have already become parabolas. Then the gravitational field has the appearance shown in the drawing above (Fig. 1). As q approaches the speed of light (the value 1), the curves open up more and more, so that the gravitational field is increasingly concentrated about the equatorial plane. Simultaneously the field strength decreases steadily toward zero, so that $\mathfrak{g}_\rho$ converges to zero as $\sqrt{(1-q^2)^3}$, and $\mathfrak{g}_z$ and u as $\sqrt{(1-q^2)}$.

45. Let us consider a body whose elementary particles are all completely at rest relative to each other, as it may be at absolute zero temperature. For this body the inertial mass is identical with the gravitational mass, let us denote it as rest mass m_0.

Now let the elementary particles in this body start to vibrate, due to an increase in temperature, for example. Let the average speed of a particle be q, then:

$$\frac{E_0}{\sqrt{1-q^2}} \sim E_0 + \frac{1}{2}E_0 q^2$$

is the average energy of a moving particle, if E_0 is its rest energy. The gravitational mass of the particle is given by (113):

$$E_0\sqrt{1-q^2} \sim E_0 - \frac{1}{2}E_0 q^2.$$

So the gravitational mass m_g of the whole body being considered decreases as the internal motion of its elementary particles increases. In fact we can estimate the change in gravitational mass if we know the magnitude of the internal motion. Let it be Q ergs, then:

$$m_g \sim m_0 - Q,$$

[48] | if the mass is specified in ergs, or:

$$m_g \sim m_0 - \frac{Q}{9\cdot 10^{20}},$$

if the mass is calculated in grams. If we impart our body a motion of velocity v, then the total velocity q' of an elementary particle, moving with velocity q *relative* to the body in a direction at angle ϑ with respect to the direction of v, becomes by the addition theorem of velocities:[20]

$$q'^2 = \frac{q^2 + v^2 + 2qv\cos\vartheta - q^2v^2\sin^2\vartheta}{(1 + q\,v\cos\vartheta)^2}.$$

The energy of this particle can be calculated:

20 M. Laue, *Das Relativitätsprinzip*. p. 43.

$$\frac{E_0}{\sqrt{1-q'^2}} = \frac{E_0(1+qv\cos\vartheta)}{\sqrt{1-q^2}\sqrt{1-v^2}}.$$

We assume that the particles in the body oscillate at random, so that each direction of q occurs equally often. Then for a large number N of particles at each moment a fraction:

$$dN = N\frac{2\pi\sin\vartheta d\vartheta}{4\pi} = \frac{1}{2}N\sin\vartheta d\vartheta$$

moves such that the direction of motion q of these dN particles makes an angle with respect to the direction of v lying between ϑ and $\vartheta + d\vartheta$. We obtain the energy of all of the N particles by multiplying the energy value of a particle, just calculated, by the number dN, and integrating over ϑ from 0 to π. Since:

$$\int_0^\pi \cos\vartheta\sin\vartheta d\vartheta = 0; \qquad \int_0^\pi \sin\vartheta d\vartheta = 2,$$

the total energy of the N particles is obtained as:

$$\frac{E_0}{\sqrt{1-q^2}\sqrt{1-v^2}} = \frac{E_0}{\sqrt{1-q^2}}\left(1+\frac{1}{2}v^2+\ldots\right).$$

| Thus the inertial mass m of a body, as is well known, simply equals its total energy [49]
content, even when its elementary particles are in vibration

$$m = \frac{E_0}{\sqrt{1-q^2}} \sim E_0 + \frac{1}{2}E_0 q^2.$$

Let us again call the energy of internal motion Q, then:

$$m \sim m_0 + Q \text{ erg}$$

or:

$$m \sim m_0 + \frac{Q}{9\cdot 10^{20}} \text{ grams}.$$

Inertial mass and gravitational mass of a body are completely identical only if the body's elementary particles execute no internal motion. Hidden motion of the elementary particles cause an increase of the inertial mass and a decrease of the gravitational mass.

Since hidden motion is certainly always present in any matter, increasing with increasing temperature, it further follows:

For any substance, the ratio of gravitational to inertial mass, and therefore also the so-called gravitational constant, is a function of the temperature, which decreases with increasing temperature.

For not excessively large thermal motion we can use the approximate values for m_g and m that we just calculated:

$$\frac{m_g}{m} = \frac{m_0 - Q}{m_0 + Q} = 1 - 2\frac{Q}{m},$$

if m and Q are both calculated in ergs, or:

$$\frac{m_g}{m} = 1 - 2\,\frac{Q}{9 \cdot 10^{20} \cdot m},$$

if Q is in ergs and m in grams; $Q:m$ is that part of the heat contents of a unit of mass of the body that represents the kinetic energy of the molecular motion.

The change of the gravitational constant with temperature is of different amounts in different materials, such that it is larger the larger the part of the material's specific heat that corresponds to the kinetic energy of molecular motion.

[50] In general the specific heat of bodies is larger the smaller the atomic weights of its constituents. Therefore the propositions that follow from our theory of gravity might be tested experimentally by determining the acceleration of gravity once with a pendulum whose bob consists of Lithium, and again under exactly the same conditions with a pendulum whose bob consists of Lead. The second pendulum should give a larger value for the acceleration of gravity than the first. To assess the feasibility of the experiment let us calculate the variation of the ratio m_g/m for that substance which must exhibit it to the greatest extent, namely hydrogen gas. It has a specific heat c_v of 2.5 cal per gram and per degree Celsius, that is converted into erg $1.05 \cdot 10^8$. This implies:

$$\frac{m_g}{m} = 1 - \frac{2.1 \cdot 10^8}{9 \cdot 10^{20}} \cdot \Theta = 1 - 2.3 \cdot 10^{-13} \cdot \Theta,$$

where Θ is the absolute temperature at which the measurement is performed. This yields at:

$$\Theta = 300^\circ \qquad 1 - 7 \cdot 10^{-11},$$
$$\Theta = 6000^\circ \qquad 1 - 14 \cdot 10^{-10}.$$

Thus it would have to be possible to determine the acceleration of gravity accurately to a fraction 10^{-11} or 10^{-12} in order to find differences for different pendulum bob materials, when observing at room temperature. Similar accuracy would be required when searching for a variability of the gravitational constant, say by astronomical means. For even though the higher temperatures of many celestial bodies could play a role here, one must also note the greater atomic weights of the materials that constitute most celestial bodies.

Gravitational mass and inertial mass are indistinguishable in practice.

Longitudinal Waves in the Aether

46. It seems that the most interesting consequence of the theory of gravitation developed here, which incidentally was also pointed out already by | M. Abraham[21] when [51]
he set up his ansatz, is the prediction of longitudinal waves in the aether. This can be immediately seen from the form of the equations (85) and (86). The equations (85) are absolutely identical with the equations of motion of a compressible and perfectly elastic fluid that executes infinitesimal, non-vortical motions, if one views ω as the velocity potential; and equation (86) is the so-called continuity equation if one can set $H = 0$, which is permitted, at least in vacuo. Here it must be supposed that $(\mathfrak{g}, iu)$ and (k, iw) are proportional to each other, which by the superposition principle is the case in vacuo. Then, for ω the wave equation results:

$$\frac{\partial^2\omega}{\partial x^2} + \frac{\partial^2\omega}{\partial y^2} + \frac{\partial^2\omega}{\partial z^2} - \frac{\partial^2\omega}{\partial t^2} = 0$$

from which it can be seen that the speed of longitudinal waves in the aether is 1, that is equal to the speed of light waves.

The analogy to longitudinal waves in a compressible fluid suggests that longitudinal waves in the aether must radiate from oscillating material particles. For, when a material particle oscillates, then (1) the sinks of the vector $\mathfrak{g}$ move back and forth periodically and (2) the amount of the sink varies periodically, since the particle's gravitational mass reaches a minimum at the time of greatest motion, and a maximum at the moment when our particle is at rest. Thus two different longitudinal spherical waves originate around an oscillating material particle, and one can already note that the second type has twice the frequency of the first.

We see from this that light waves, if emitted by oscillating electrons, and x-rays, which originate upon sudden deceleration of moving electrons, must always be accompanied by radiation of longitudinal waves. However, the same cannot be immediately said of such light waves that consist of exploded | dipoles (as in section 32). [52]
Let us now calculate how great the intensity of gravitational waves emitted by an oscillating material particle is.

For simplicity we will assume that the vectors $\mathfrak{h}$ and $\mathfrak{v}$ are zero in the whole surrounding of the particle when it is at rest, and also that $u = 0$. Then the energy of the particle at rest is always:

$$E = \int H dV.$$

If this particle is at an equilibrium position, then in particular:

$$E_0 = \int H_0 dV.$$

21 M. Abraham, *Physik. Zeitschr.* 13. p. 1. 1912.

At other positions we have $E > E_0$, and to calculate E we may have to take into account the further surroundings of the particle.

As the particle moves through its equilibrium position with speed q, its energy is:

$$E = \frac{E_0}{\sqrt{1-q^2}},$$

but from (69):

$$\int H dV = E_0\sqrt{1-q^2} = E(1-q^2).$$

If E is the total energy of oscillation, which stays constant during the oscillation (apart from radiation damping), then at the moment of maximum amplitude, when $q = 0$:

$$\int H dV = E,$$

and at the moment of crossing the equilibrium position, when q reaches its maximum:

$$\int H dV = E(1-q^2).$$

We will now make use of the following abbreviations:

$$\left.\begin{aligned} \int H dV &= \mu, \\ E &= \mu_0. \end{aligned}\right\} \tag{127}$$

[53] | If we think of μ as the massive particle, we can visualize the whole oscillation process thus: (1) the center of mass of μ moves with a periodic velocity, which we will call q and (2) concurrently the mass of the particle μ changes periodically; we can calculate to sufficient accuracy:

$$\mu = \mu_0(1-q^2), \tag{128}$$

where μ_0 is the value for $q = 0$. For the intensity of the emitted waves it will be inessential whether, in doing this, the particle also undergoes any kind of change of shape.

In the following let us confine attention to the case of linear oscillation of the particle, where therefore q does not change direction. We will put:

$$q = a \sin 2\pi n t. \tag{129}$$

In place of the vector of gravity $\mathfrak{g}$ and u we first calculate quantities defined somewhat differently, which we will call $\mathfrak{g}'$ and u', and which shall satisfy the following equations:

$$\left.\begin{aligned} \mathfrak{g}' &= \nabla\omega', \quad u' = -\frac{\partial\omega'}{\partial t}, \\ \operatorname{div}\mathfrak{g}' &+ \frac{\partial u'}{\partial t} = -\gamma H, \end{aligned}\right\} \tag{130}$$

In vacuo, where $\mathfrak{g}$ and k, and u and w are identical to each other, the definition of $\mathfrak{g}'$ and u' as well as ω' agrees with that of $\mathfrak{g}$, u and ω, but not in the interior of the material particle, because there k and $\mathfrak{g}$, and u and w may not be regarded as identical. From (130) one derives the well-known differential equation for ω':

$$\frac{\partial^2\omega'}{\partial x^2} + \frac{\partial^2\omega'}{\partial y^2} + \frac{\partial^2\omega'}{\partial z^2} - \frac{\partial^2\omega'}{\partial t^2} = -\gamma H, \tag{131}$$

which agrees completely with the differential equation obeyed by the scalar potential of the electric field about a moving charged particle, if $+\gamma H$ is taken as the density of electric charge. But this equation can be easily integrated by the method given by E. Wiechert[22] if one confines attention to | regions of space whose distances from the [54] center of mass of the moving particle are infinitely large compared to the particle's dimensions. Namely:

$$\omega'(x, y, z, t) = \frac{\gamma}{4\pi}\left(\frac{\mu}{r\,(1 - q\cos\vartheta)}\right)_{x_1, y_1, z_1, t_1}, \tag{132}$$

where:

$$\mu = \int H dV \qquad \text{at time } t_1.$$

Here x_1, y_1, z_1 denotes that point on the orbit described by the particle's center of mass from which a light wave would just arrive at the point in space (x, y, z) under consideration at the time t, and t_1 is the moment at which the particle's center of mass is just passing (x_1, y_1, z_1). r is the connecting segment from (x_1, y_1, z_1) to (x, y, z). Since we have put the speed of light equal to unity we have:

$$t_1 = t - r.$$

Let ϑ be the angle between the radius vector r, directed from (x_1, y_1, z_1) to (x, y, z), and the direction of motion q. According to (129) q is given as a function of t_1: $q = a\sin 2\pi n \cdot t_1$ (t now has a different meaning). If we take the direction of motion of the oscillating particle as the direction of the z-axis, then x_1, y_1 are constants and z_1 is a function of t_1:

22 E. Wiechert, *Ann. d. Phys.* 4. p. 682. 1901.

$$z_1 = \mathfrak{f}(t_1), \quad \frac{dz_1}{dt_1} = q = \mathfrak{f}'(t_1).$$

According to (129) we have to set:

$$z_1 = -\frac{a}{2\pi n}\cos 2\pi n t_1, \quad \frac{dz_1}{dt_1} = q = a \sin 2\pi n t_1.$$

The equation:

$$t_1 = t - r = t - \sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2},$$

with the constant values for x_1 and y_1 and $z_1 = \mathfrak{f}(t_1)$ substituted, determines the quantity t_1 implicitly as function of (x, y, z, t), and the differential quotients

$$\frac{\partial t_1}{\partial x}, \quad \frac{\partial t_1}{\partial y}, \quad \frac{\partial t_1}{\partial z}, \quad \frac{\partial t_1}{\partial t}$$

[55] | can be calculated without effort. Since q and therefore also μ are functions of t_1, they can also be differentiated with respect to x, y, z, t, where we will use the abbreviation:

$$\frac{dq}{dt_1} = \dot{q}, \quad \frac{d\mu}{dt_1} = \dot{\mu}.$$

Finally one can also perform the differentiation of r and $rq\cos\vartheta = (z - z_1)q$ and calculate now the vector $\mathfrak{g}'$ and u':

$$\mathfrak{g}_x' = \frac{\partial \omega'}{\partial x}, \quad \mathfrak{g}_y' = \frac{\partial \omega'}{\partial y}, \quad \mathfrak{g}_z' = \frac{\partial \omega'}{\partial z}, \quad u' = -\frac{\partial \omega'}{\partial t}.$$

The calculations have been carried through exactly by M. Abraham in his *Theorie der Elektrizität* vol. II in § 13 p. 92ff, therefore I need only write down the results. Because everything is symmetric about the z-axis I will give only two components of $\mathfrak{g}'$: namely $\mathfrak{g}_z'$ and $\mathfrak{g}_\rho' \perp \mathfrak{g}_z'$:

$$\left.\begin{aligned}
\mathfrak{g}_\rho' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(1-q^2)\sin\vartheta - \frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3}\dot{q}\cos\vartheta\sin\vartheta \\
&\qquad - \frac{\gamma}{4\pi r(1-q\cos\vartheta)^2}\dot{\mu}\sin\vartheta, \\
\mathfrak{g}_z' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(\cos\vartheta - q) - \frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3}\dot{q}\cos^2\vartheta \\
&\qquad - \frac{\gamma}{4\pi r(1-q\cos\vartheta)^2}\dot{\mu}\cos\vartheta, \\
u' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}q(\cos\vartheta - q) - \frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3}\dot{q}\cos\vartheta \\
&\qquad - \frac{\gamma}{4\pi r(1-q\cos\vartheta)^2}\dot{\mu},
\end{aligned}\right\} \quad (133)$$

These expressions decompose into two parts, the first being proportional to r^{-2}, and the second to r^{-1}. The two summands of the second part contain as a factor either:

$$\dot{q} = 2\pi n a\cos 2\pi nt = \frac{2\pi}{\lambda}a\cos 2\pi nt$$

or:

$$\mu = -\mu_0 q\dot{q} = -\frac{2\pi}{\lambda}\mu_0 a^2\sin 2\pi nt\cos 2\pi nt$$

| The second part of the expressions is related to the first in order of magnitude as [56]
$1/\lambda : 1/r$, where λ is the wavelength of light corresponding to the wave number n.

Let us confine attention to those wave numbers n whose wavelength of light λ is infinitely large in comparison with the dimensions of the particle, in the sense *that there are distances r from the particle that are infinitely small compared to* λ, *but still infinitely large compared to the dimensions of the particle*.

By this assumption there are values of r for which the formulas (133) are still valid, although r is infinitesimal compared to λ. For these *small values of r* we can calculate to good approximation:

$$\left.\begin{aligned}
\mathfrak{g}_\rho' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(1-q^2)\sin\vartheta, \\
\mathfrak{g}_z' &= -\frac{\gamma\mu}{4\pi r^2(1-q\cos\vartheta)^3}(\cos\vartheta - q), \\
u' &= -\frac{\gamma\cdot\mu}{4\pi r^2(1-q\cos\vartheta)^3}q(\cos\vartheta - q).
\end{aligned}\right\} \quad (134)$$

It is easy to show that these formulas agree completely with the formulas (123), which represent the field of a massive particle moving with speed q.

To see this we have to note that r stands for the distance of the point (x, y, z) from the position of the particle at the time $t_1 = t - r$. If we denote the distance between (x, y, z) and the position of the particle at time t by r', and the angle between r' and the z-axis by ϑ', we see immediately from Fig. 2 that:

$$r \sin\vartheta = r' \sin\vartheta',$$
$$r \cos\vartheta = r' \cos\vartheta' + q(t - t_1) = r' \cos\vartheta' + rq.$$

An elementary calculation leads to the formula:

$$r'^2(1 - q^2 \sin^2\vartheta') = r^2(1 - q\cos\vartheta)^2$$

or:

$$r' \sqrt{\left(1 + \frac{q^2}{1-q^2}\cos^2\vartheta'\right)}\sqrt{1-q^2} = r(1 - q\cos\vartheta).$$

[57] | Noting that:

$$\sin\vartheta = \frac{r'}{r}\sin\vartheta',$$

$$\cos\vartheta - q = \frac{r'}{r}\cos\vartheta'$$

and putting as before (121):

$$\sqrt{1 + \frac{q^2}{1-q^2}\cos^2\vartheta'} = p,$$

we readily derive from (134) the formulas (125) where:

$$m_0 = \frac{\mu}{\sqrt{1-q^2}}$$

in agreement with formula (113).

At distances that are large compared to the dimensions of the particle, but which still belong to its closer vicinity (r small compared to λ) the vector $\mathfrak{g}'$ and u' agrees completely with the vector of gravity $\mathfrak{g}$ and u of the moving massive particle.

Hence it follows that this is valid for all distances that are large compared to λ. Only in the interior of the particle and in the nearest vicinity can $\mathfrak{g}', u'$ be distinguished from $\mathfrak{g}, u$. But since these regions do not interest us I will simply omit the primes in the following and put:

$$\mathfrak{g}' = \mathfrak{g}, \qquad u' = u.$$

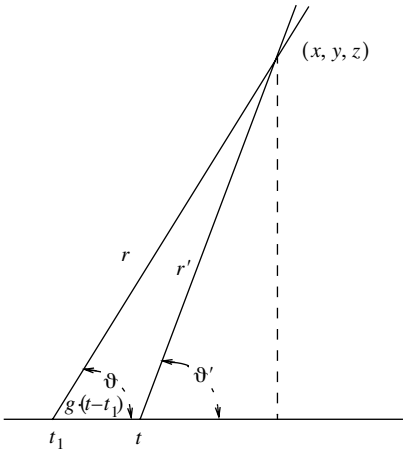

Figure 2

| At large distances, where r is infinitely large compared to λ, we have the pure longitudinal spherical waves that we desire. We need to take into account only the second part of the expressions (133): [58]

$$\mathfrak{g}_\rho = -\frac{\gamma\mu}{4\pi r}\left\{\frac{\dot{q}}{(1-q\cos\vartheta)^3}\cos\vartheta\sin\vartheta - \frac{q\dot{q}}{(1-q\cos\vartheta)^2(1-q)^2}\frac{\sin\vartheta}{}\right\}$$

by noting that:

$$\dot{\mu} = -\mu_0 q\dot{q} = -\frac{\mu}{1-q^2}q\dot{q}.$$

A minor transformation yields:

$$\mathfrak{g}_\rho = -\frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q})\sin\vartheta.$$

If we transform $\mathfrak{g}_z$ and u similarly, we finally obtain the following expressions for the variables of state in the longitudinal wave *at large distances from the oscillating particle*:

$$\left.\begin{aligned} \mathfrak{g}_\rho &= -\frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q})\sin\vartheta, \\ \mathfrak{g}_z &= -\frac{\gamma\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q})\cos\vartheta, \\ u &= -\frac{\gamma\cdot\mu}{4\pi r(1-q\cos\vartheta)^3(1-q^2)}(\dot{q}\cos\vartheta - q\dot{q}). \end{aligned}\right\} \tag{135}$$

This shows that $\mathfrak{g}$ is always oriented radially toward the particle, regarding the norm $\mathfrak{g} = u$. Furthermore the two waves mentioned above are clearly recognizable; the term $\dot{q}\cos\vartheta$ is due to the back and forth motion of the sink of the vector $\mathfrak{g}$, and the term $q\dot{q}$ is due to the change of the particle's gravitational mass as it is moving. Since:

$$q\dot{q} = a^2 \sin 2\pi nt \cos 2\pi nt = \frac{1}{2}a^2 \sin 4\pi nt$$

the second wave has twice the frequency of the first. If the particle's displacements are very large, so that q is not much less than 1, then the factor before the parenthesis causes the emitted radiation to consist of oscillations that are not pure sine waves. There is little interest for us to discuss this complication more precisely here, because motions of material particles with velocity not *far* below 1 are well known to be
[59] extremely rare. | Therefore we want to discuss exclusively the case that q is very small. Then $q\cdot\dot{q}$ can also be dropped compared to $\dot{q}\cos\vartheta$, and we retain only the first wave:

$$\left.\begin{aligned} \mathfrak{g} &= -\frac{\gamma\mu}{4\pi r}\dot{q}\cos\vartheta, \\ u &= -\frac{\gamma\mu}{4\pi r}\dot{q}\cos\vartheta. \end{aligned}\right\} \tag{136}$$

Here μ is simply the *mass* of the particle, since for small velocities inertial and gravitational mass are identical. We substitute in (136):

$$\dot{q} = \frac{2\pi}{\lambda}a\cos 2\pi nt$$

and calculate the energy current of the wave in direction ϑ to be:

$$\mathfrak{g}\,u = \frac{\gamma^2\mu^2}{16\pi^2 r^2}\left(\frac{2\pi}{\lambda}\right)^2 a^2\cos^2\vartheta\cos^2 2\pi nt$$

from which one obtains the intensity by integration over a unit of time:

$$J_g = \frac{\gamma^2\mu^2}{16\pi^2 r^2}\left(\frac{2\pi}{\lambda}\right)^2\frac{1}{2}a^2\cos^2\vartheta. \tag{137}$$

Let us now consider a very large number N of oscillating particles, which are oriented quite at random in space, but which all oscillate linearly with amplitude a. Since in the spherical zone between ϑ and $\vartheta + d\vartheta$ their number is:

$$dN = \frac{N}{2}\sin\vartheta d\vartheta$$

we obtain the whole intensity by multiplying (137) by this value of dN and integrating over ϑ, with the result obtained in this way:

$$NJ_g = \frac{1}{6}N\frac{\gamma^2\mu^2}{16\pi^2 r^2}\left(\frac{2\pi}{\lambda}\right)^2 a^2. \tag{138}$$

Let us compare this value with the corresponding value of the intensity of the emitted light. We consider a material particle with charge ε whose direction of oscillation makes an angle ϑ with the ray direction, and which oscillates with speed $q = a\sin 2\pi nt$. | The intensity of the light emitted by this particle is, in the system [60] of units used by us:[23]

$$J_e = \frac{\varepsilon^2}{16\pi^2 r^2}\left(\frac{2\pi}{\lambda}\right)^2\frac{1}{2}a^2\sin^2\vartheta \tag{139}$$

and:

$$NJ_e = \frac{1}{3}N\frac{\varepsilon^2}{16\pi^2 r^2}\left(\frac{2\pi}{\lambda}\right)^2 a^2. \tag{140}$$

Accordingly the ratio of the intensities for the electric and gravitational waves radiated by the same particle is:

$$\frac{\bar{J}_e}{\bar{J}_g} = 2\frac{\varepsilon^2}{\gamma^2\mu^2}. \tag{141}$$

In this formula ε means the amount of charge of the particle, calculated in a unit that is the $\sqrt{4\pi}^{\text{th}}$ part of the usual electrostatic unit, so that:

$$\varepsilon^2 = 4\pi\varepsilon_s^{\,2},$$

if ε_s is the charge calculated in the ordinary electrostatic unit. Further, μ means the mass calculated in ergs, that is:

$$\mu = c^2 m,$$

where c denotes the speed of light and m the mass in grams. Finally, if we denote the charge in electromagnetic units by e, then:

23 Cf. E. Wiechert, *Ann. d. Phys.* 4. p. 688. 1901.

$$\frac{\varepsilon_s}{c} = e.$$

Thus we obtain:

$$\frac{\bar{J}_e}{\bar{J}_g} = 2\frac{4\pi}{\gamma^2 c^2}\left(\frac{e}{m}\right)^2,$$

or, since by (113) the ordinarily so-called gravitational constant κ has the value:

$$\kappa = \frac{\gamma^2 c^4}{4\pi},$$

we have:

$$\frac{\bar{J}_e}{\bar{J}_g} = 2\frac{c^2}{\kappa}\left(\frac{e}{m}\right)^2, \tag{142}$$

$$c = 3 \cdot 10^{10}, \quad \kappa = 6,65 \cdot 10^{-8}, \quad e/m = 1,75 \cdot 10^{7},$$

[61] | so the result for the radiation emitted by oscillating electrons is:

$$\frac{\bar{J}_e}{\bar{J}_g} = 8,3 \cdot 10^{42}.$$

To appreciate this number properly let us take the square root, $\sim 3 \cdot 10^{21}$. Then we see the following: the intensity of gravity radiation emitted by a radiating point at a distance of 1 cm is just as intense as the radiation of light emitted by it at the distance of $3 \cdot 10^{21}$ cm, that is a distance a hundred million times the diameter of the Earth's orbit, or about 3000 light years. Here this ratio is the same for all wavelengths.

The gravitational radiation emitted by oscillating electrons (or by any oscillating mass particle) is so extraordinarily weak that it is unthinkable ever to detect it by any means whatsoever.

This makes it transparent why the longitudinal radiation of the aether apparently plays no role whatsoever in the balance of nature. It would probably be very imprudent to claim that the longitudinal waves, which certainly as such are possible at any amplitude, could not arise in appreciable intensity for other than oscillatory processes of material particles. We can only claim this much with certainty, that these processes would have to be of a highly peculiar type. So if one could ever prove the existence of gravitational waves, the processes responsible for their generation would probably be much more curious and interesting than even the waves themselves.

The Theorem on the Relativity of the Gravitational Potential

47. The theory advocated in this work differs from the theory generally adopted to date mainly because it must put the real vacua in contrast with yet another, ideal vacuum, as in the theory of gases the real gases vs. the ideal gas. For in real vacua, due to

the proximity of material bodies, traces of the states $\rho, \mathfrak{v}$ are always present and H is not absolutely zero; therefore the law of superposition of | the states of the aether, [62] which characterizes the absolute vacuum, is always valid only as a good approximation, admittedly to such a good approximation that one can hardly hope very easily to substantiate deviations from this law. However it may nevertheless be possible, in very strong electric fields where $\mathfrak{e}$ and φ are large,[24] or in very strong magnetic fields, to make observations which contradict our present-day ideas about the vacuum, and such observations should be viewed as the greatest encouragement on the path followed by me.

Such observations would concern a vacuum that already deviates rather strongly from the ideal vacuum. But among the state variables is one which appears to influence processes even in a vacuum that otherwise deserves to be called almost ideal, and that is the gravitational potential ω. If the quantities $\mathfrak{v}$ and ρ, as well as H, are so small that one can no longer speak of any noticeable influence on physical laws, but if in this good vacuum ω still has a large value, then we do not have $\mathfrak{e} = \mathfrak{d}, \mathfrak{b} = \mathfrak{h}$ but, as shown by (93) and (94):

$$\mathfrak{e} = e^{-\gamma\omega} \cdot \mathfrak{d}, \quad \mathfrak{b} = e^{-\gamma\omega} \cdot \mathfrak{h}.$$

From this it is evident that the dielectric constant of the vacuum is no longer 1 in a region where the gravitational potential ω is present, instead $K = e^{+\gamma\omega}$, similarly the permeability of the vacuum is no longer 1, but $M = e^{-\gamma\omega}$. But the product of the two is again $K \cdot \mathrm{M} = 1$, that is *the speed of light in a region at gravitational potential* ω *is the same as in a region with zero gravitational potential*.

But we can go much further. Let the mean value of the gravitational potential in our region be ω_0, an arbitrarily large but constant value. Then we can decompose the gravitational potential, which | is not constant due to the presence of matter and of [63] gravitational fields, into two parts:

$$\omega = \omega_0 + \omega_1.$$

The second, variable term takes on large values at most in the interior of material bodies present in the region, in vacuo itself ω_1 is small. If we define:

$$H_1 = e^{+\gamma\omega_0} H = e^{-\gamma\omega_1} H'(\mathfrak{d}, \mathfrak{h}, \rho, \mathfrak{v}, \mathfrak{g}, u),$$

$$\mathfrak{e}_1 = \frac{\partial H_1}{\partial \mathfrak{d}}, \quad \mathfrak{b}_1 = -\frac{\partial H_1}{\partial \mathfrak{h}}, \quad \varphi_1 = -\frac{\partial H_1}{\partial \rho}, \quad \mathfrak{f}_1 = \frac{\partial H_1}{\partial \mathfrak{v}},$$

$$\mathfrak{k}_1 = \frac{\partial H_1}{\partial \mathfrak{g}}, \quad w_1 = -\frac{\partial H_1}{\partial u}, \quad \frac{\partial H_1}{\partial \omega_1} = -\gamma H_1,$$

24 Cf. II. p. 24 [in chap. 4 of the original, which is not included in this translation].

then in the region at gravitational potential ω_0 exactly the same equations hold for the quantities denoted with the index 1 as in a region at gravitational potential 0 for the quantities without index. From this it follows directly:

If two empty regions differ only in this, that the gravitational potential in one of them has a very large average value ω_0, and in the other one it has the average value zero, then this does not have the least influence on the size and form of the electrons and other material elementary particles, on their charges, their laws of oscillation and other laws of motion, on the speed of light, and in general on all physical relations and processes.

Through this theorem, which could be called the principle of the relativity of the gravitational potential, the theory of gravitation developed by me differs in principle and most sharply from the theories of A. Einstein and M. Abraham. I share the view of the latter that if this theorem were not valid, it would mean the demise of the entire principle of relativity. On the other hand I believe to have shown that the postulates I assumed lead nowhere to contradictions with experience, that in particular according to my theory only imperceptibly small deviations are to be expected from the law of proportionality of gravitational mass and inertial mass. |

[64] *Concluding Remarks*

48. Above I believe to have pursued the general theory of matter as far as is possible today. The next progress must occur through experiments, and therefore I want to discuss briefly what possibilities offer themselves to experiment.

Gravity, the preparation of whose experimental investigation was the main goal that was on my mind in this research, shows itself as obstinate as ever. It was possible to implement the theory of gravitation completely so that it is in accord with the principle of relativity as well as with all empirical facts known about gravity, and it also yields two new results that seem extremely interesting on first sight. But a closer look shows that these theoretical results provide no prospects for a successful experiment. The first result is that the ratio of gravitational to inertial mass depends on temperature, and that the dependence on temperature is much stronger for bodies of small atomic weight than for bodies of large atomic weight. Because the differences to be expected from theory for the acceleration of gravity of different substances is of the order 10^{-12} to 10^{-11}, it is experimentally useless. The second result is that there must be longitudinal waves in the aether, for which it may be worth searching. Of the processes known to us, the oscillations of atoms and electrons are relevant, which must produce longitudinal gravitational waves as they produce light. However, for electron oscillations the intensity of gravitational waves is to that of the light waves of any frequency as $1{:}8.3 \cdot 10^{42}$, and we must therefore deem the existence of any reagent that would react to them as totally ruled out. Thus no way can be given to search for these longitudinal waves, which by themselves are certainly highly interesting.

The next thing immediately suggested by the theory would be an investigation whether there are to be found, in very strong electric or magnetic | fields or in field- [65]
free regions at a very high potential, any deviations from the laws of Maxwell that are valid in the ideal vacuum. These would be high precision measurements, to be performed according to a theoretically well thought-out plan. Of course it is doubtful whether this will lead to success; but if there were positive results, they would give important hints to the theory how the next steps should be taken.

In somewhat loose connection with the remaining theory is the conception laid out by me in the sections 31 to 36 about the quanta of action and the light of band spectra. The conception is very vague and full of hypotheses, nevertheless I believe that one could draw some conclusions from it which should give rise to new spectroscopic investigations.

EDITORIAL NOTE

[1] In the original, the second equation is misprinted as $\mathfrak{g}_y = \dfrac{\partial \omega}{\partial x}$.

GUSTAV MIE

REMARKS CONCERNING EINSTEIN'S THEORY OF GRAVITATION

Originally published as "Bemerkungen zu der Einsteinschen Gravitationstheorie" in Physikalische Zeitschrift 15, 1914, 3: 115–122 (received December 28, 1913) and 4: 169–176 (received 28 December 1913). Author's note: Greifswald, Physical Institute, December 24, 1913.

1. Introduction
2. General Theory of Gravitation with a Tensor Potential
3. Impossibility of the Identity of Gravitational and Inertial Mass
4. Special Assumptions of Einstein's Theory
5. The Fundamental Equations of Einstein's Theory
6. The Energy Tensor
7. Theorem of the Relativity of the Gravitational Potential
8. Equality of Inertial and Gravitational Mass of Closed Systems in the Theories of Einstein and Mie
9. Internal Contradiction in Einstein's Auxiliary Assumptions
10. Appendix: Nordström's Two Theories of Gravitation
11. Conclusion: Summary

1. INTRODUCTION

In his interesting paper *Outline of a Generalized Theory of Relativity* etc.[1] Mr. Einstein says that by introducing a variable speed of light he has broken out of "the confines of the theory known at present as the theory of relativity," and also on other occasions he frequently contrasts his theory with the "old theory of relativity." To someone who immerses himself sufficiently deeply in the development it may become clear in what respect Mr. Einstein can speak of a new relativity; nevertheless, one may suppose that the cited passages may lead a more cursory reader to the wrong view, that one is really dealing here with a break with the theory of relativity as presently known. It is therefore certainly not without interest to present below, using

1 A. Einstein and M. Grossmann, *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Published as an offprint by B. G. Teubner, 1913.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

methods recently developed by me in a larger work about the theory of matter,[2] which strictly followed Minkowski's concept of the principle of relativity, a general theory of gravitation with a tensor potential, which includes Einstein's theory as a special case. Contrary to the rather inscrutable formulas of Einstein, the methods I use have the advantage that they yield clearly transparent expressions. In this fashion, it then becomes possible to comprehend the nature of Einstein's theory better, and in particular to clarify the so-called generalization of the principle of relativity, and furthermore to compare it with the theory of gravitation suggested by myself[3] and with
[116] that | of Nordström,[4] which deviates only slightly from mine.

2. GENERAL THEORY OF GRAVITATION WITH A TENSOR POTENTIAL

The essential difference between Einstein's theory of gravitation and my own is that in the former the gravitational field is described not by means of a four-vector, but by means of a spacetime quantity of third rank, which is related to a tensor (that is, a quantity of rank two) in the same way as a four-vector (a quantity of rank one) is related to a scalar (a quantity of rank zero). Because a tensor has 10 components, the gravitational field of Einstein has 40 components, which can be easily surveyed if each quadruple is associated with one component of a tensor, similar to associating the four components of a four-vector with a scalar. To make it intuitive I take the liberty of calling such a quantity a vector of tensors. I shall denote all the spatial components of this vector of tensors by $\mathfrak{g}$, and the temporal components by iu (cf. *Theory of Matter* III, p. 28). The four components that belong to the tensor component numbered by (μ, ν) are therefore: $\mathfrak{g}_{\mu\nu x}, \mathfrak{g}_{\mu\nu y}, \mathfrak{g}_{\mu\nu z}, iu_{\mu\nu}$ The indices μ, ν have to run over the values 1 to 4, where we have $\mathfrak{g}_{\mu\nu} = \mathfrak{g}_{\nu\mu}$, $u_{\mu\nu} = u_{\nu\mu}$. In addition to the vector of tensors $\mathfrak{g}_{\mu\nu}, iu_{\mu\nu}$ I introduce a second one, which I will denote by the letters $\mathfrak{k}$ and w (loc. cit. III, p. 28), that is $(\mathfrak{k}_{\mu\nu}, iw_{\mu\nu})$. This second vector of tensors shall describe the gravitational field equally well as the first; the two shall be related to each other in a similar way as the electric field strength is to the electric displacement, or the elastic stress is to the elastic strain. In an ideal vacuum, that is, at infinite distance from matter, they shall be equal, $(\mathfrak{k}_{\mu\nu}, iw_{\mu\nu}) = (\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$. In addition to the vector of tensors, for which we may choose any one of the two just named, the complete description of the state of the aether in a gravitational field requires another four dimensional tensor quantity, which I will denote by $\omega_{\mu\nu}$, and which one may call the gravitational potential. The denseness[1] of gravitational mass must be a tensor quantity as well in this theory, I will call its components $h_{\mu\nu}$ and for now make no further assumptions about how they are calculated from the state variables characterizing the

2 G. Mie, *Ann. d. Phys.*, Abhandlung I: 37, 511, 1912; Abhandlung II: 39, 1, 1912; Abhandlung III: 40, 1, 1913 [selections from I and III are included in this volume].

3 G. Mie, l. c. III, p. 25 ff.

4 G. Nordström, *Physik. Zeitschr.* 13, 1126, 1912; *Ann. d. Phys.* 40, 856, 1913. In the meantime Mr. Nordström has put up a second, quite different theory, which I shall briefly discuss at the end of the present paper.

material body. I now put down the following 50 equations for the 50 quantities $\mathfrak{g}_{\mu\nu}, i \cdot u_{\mu\nu}, \omega_{\mu\nu}$, by which the gravitational field is completely described (cf. l. c. III, p. 28):

$$\left.\begin{aligned} \mathfrak{g}_{\mu\nu x} &= \frac{\partial \omega_{\mu\nu}}{\partial x} \\ \mathfrak{g}_{\mu\nu y} &= \frac{\partial \omega_{\mu\nu}}{\partial y} \\ \mathfrak{g}_{\mu\nu z} &= \frac{\partial \omega_{\mu\nu}}{\partial z} \\ u_{\mu\nu} &= -\frac{\partial \omega_{\mu\nu}}{\partial t} \end{aligned}\right\} \tag{1}$$

$$\frac{\partial \mathfrak{f}_{\mu\nu x}}{\partial x} + \frac{\partial \mathfrak{f}_{\mu\nu y}}{\partial y} + \frac{\partial \mathfrak{f}_{\mu\nu z}}{\partial z} + \frac{\partial w_{\mu\nu}}{\partial t} = -\kappa \cdot h_{\mu\nu}. \tag{2}$$

The quantity κ is a universal constant, which is denoted in the same way by Einstein, whereas I have used the letter γ for it in the *Theory of Matter*. Because eqs. (1) and (2) admit Lorentz transformations, the principle of relativity is satisfied in this theory. Incidentally, the eqs. (1) are equivalent to the following:

$$\left.\begin{aligned} \frac{\partial \mathfrak{g}_{\mu\nu x}}{\partial t} + \frac{\partial u_{\mu\nu}}{\partial x} &= 0 \\ \frac{\partial \mathfrak{g}_{\mu\nu y}}{\partial t} + \frac{\partial u_{\mu\nu}}{\partial y} &= 0 \\ \frac{\partial \mathfrak{g}_{\mu\nu z}}{\partial t} + \frac{\partial u_{\mu\nu}}{\partial z} &= 0 \end{aligned}\right\} \tag{3}$$

$$\frac{\partial \omega_{\mu\nu}}{\partial t} + u_{\mu\nu} = 0. \tag{4}$$

Multiplying eqs. (2) by $u_{\mu\nu}$, and eqs. (3) by $\mathfrak{f}_{\mu\nu x}, \mathfrak{f}_{\mu\nu y}, \mathfrak{f}_{\mu\nu z}$, and then adding it all, using (4), we have:

$$\operatorname{div} \sum u_{\mu\nu} \cdot \mathfrak{f}_{\mu\nu} + \sum \mathfrak{f}_{\mu\nu} \cdot \frac{\partial \mathfrak{g}_{\mu\nu}}{\partial t} + \sum u_{\mu\nu} \cdot \frac{\partial w_{\mu\nu}}{\partial t} - \kappa \cdot \sum h_{\mu\nu} \cdot \frac{\partial \omega_{\mu\nu}}{\partial t} = 0.$$

The summation symbols here are to be interpreted as summing over μ and ν from 1 to 4, as if the quantities numbered by μ, ν were different from those numbered by ν, μ. The sum $\Sigma u_{\mu\nu} \cdot \mathfrak{f}_{\mu\nu}$ yields an ordinary three dimensional vector that may be regarded as the energy flux (cf. loc. cit. III, p. 29). It is then easy to show that the energy principle is satisfied if there is a four dimensional scalar H, a function of all the quantities that determine the state of the aether, including also the quantities $\mathfrak{g}_{\mu\nu}, u_{\mu\nu}, \omega_{\mu\nu}$,

from which the other vector of tensors $(\mathfrak{k}_{\mu\nu}, iw_{\mu\nu})$ for the gravitational field can be
[117] derived as follows: |

$$\mathfrak{k}_{\mu\nu x} = \frac{\partial H}{\partial \mathfrak{g}_{\mu\nu x}}\,, \qquad \mathfrak{k}_{\mu\nu y} = \frac{\partial H}{\partial \mathfrak{g}_{\mu\nu y}}\,, \qquad \mathfrak{k}_{\mu\nu z} = \frac{\partial H}{\partial \mathfrak{g}_{\mu\nu z}}\,, \qquad w_{\mu\nu} = -\frac{\partial H}{\partial u_{\mu\nu}} \tag{5}$$

and which in addition satisfies the differential equation

$$\frac{\partial H}{\partial \omega_{\mu\nu}} = -\kappa \cdot h_{\mu\nu}. \tag{6}$$

The proof of this assertion is extraordinarily simple; it proceeds in the same way as that for the corresponding theorem in the theory of gravitation with a scalar potential (loc. cit. III, p. 29, 30) and therefore there is no need to write it down here once again.

Incidentally, in all differentiations in (5) and (6) the quantities with the index (μ, ν) should be regarded as formally different from those with the index (ν, μ).

The quantity H is nothing other than the rest density of energy (or, equivalently, the rest density of inertial mass); in the general theory of matter it plays the role of the Hamiltonian function per unit volume, thus, for brevity, I usually call it the Hamiltonian function.

Now we want to calculate the force experienced by a mass particle in an external gravitational field $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$. The calculation for the gravitational field with tensor potential proceeds in exactly the same way as for the gravitational field with scalar potential (loc. cit. III, p. 35–40). For the calculation we presuppose that in the space surrounding the volume occupied by the mass particle the ratio of the vector $(\mathfrak{k}_{\mu\nu}, iw_{\mu\nu})$ to the vector $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$ may be regarded as constant. For the gravitational mass of the particle the calculation yields a quantity with the following 10 components:

$$m^g_{\mu\nu} = \int_V h_{\mu\nu} dV. \tag{7}$$

The integral should range over the entire volume V occupied by the mass particle. The force acting on the particle is calculated from the double sum:

$$\mathfrak{P} = \kappa \sum_{\nu=1}^{4} \sum_{\mu=1}^{4} \mathfrak{g}_{\mu\nu} m^g_{\mu\nu}. \tag{8}$$

3. IMPOSSIBILITY OF THE IDENTITY OF GRAVITATIONAL AND INERTIAL MASS

The density of inertial mass is identical with the density of energy, which is the (4,4) component of a symmetric tensor that I will, for brevity, call the energy tensor; there-

fore one can speak about identity or unity of essence[2] of the two masses only if the gravitational mass also occurs in the form of a tensor, a tensor that would have to be completely identical to the energy tensor. We have just learned in general terms the form of the basic equations for a theory in which the denseness of gravitational mass is a tensor: $(h_{\mu\nu})$. If we now denote the components of the energy tensor by $T_{\mu\nu}$, then *the principle of the identity of gravitational and inertial mass* is:

$$h_{\mu\nu} = T_{\mu\nu}. \tag{9}$$

If this principle were satisfied, then according to *Laue's theorem*[5] all components $m^g_{\mu\nu}$ of the gravitational mass of a material body at rest (a completely stationary system) would vanish, except for m^g_{44}, and, if I call the inertial mass of the body m_i (inertia), we would have $m^g_{44} = m_i$. The gravitational force acting on the body in the gravitational field $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$ would then be:

$$\mathfrak{P} = \kappa \mathfrak{g}_{44} m_i.$$

According to this, in one and the same gravitational field the forces of gravity acting on different bodies, which are in the same state of motion, would be strictly proportional to their inertial masses.

However, it is easy to show that the identity $h_{\mu\nu} = T_{\mu\nu}$ required by the principle as discussed is impossible. I will show that no function H of $\omega_{\mu\nu}$ can be found that satisfies the differential eqs. (6) in the form required by the principle of identity of the two masses:

$$\frac{\partial H}{\partial \omega_{\mu\nu}} = -\kappa T_{\mu\nu}. \tag{10}$$

If eq. (10) were satisfied, then the following would also be valid:

$$\frac{\partial T_{\mu\nu}}{\partial \omega_{\kappa\lambda}} = \frac{\partial T_{\kappa\lambda}}{\partial \omega_{\mu\nu}}. \tag{11}$$

Now I have shown in my paper on the theory of matter (Part III, p. 34, eq. (105)) that one can very simply write down the energy tensor by means of a quantity Φ, which like H is a function of the state of the aether at the spacetime point under consideration. Of course, when calculating with Φ one has to choose different quantities to describe the state of the aether than when calculating with H. Without regard to their physical meaning, the correctly chosen quantities will be ordered simply according to their association with the four coordinate axes and denoted | by [118]
$x_1, x_2, x_3, x_4;\ y_1, y_2, y_3, y_4;\ z_1, z_2, z_3, z_4$ etc. The components of the energy tensor can then be represented by the following schema:

5 M. v. Laue, *Das Relativitätsprinzip*, 2. ed., p. 209. Published by Vieweg & Sohn. Braunschweig 1913.

$$\left.\begin{aligned} T_{11} &= \Phi - \frac{\partial \Phi}{\partial x_1} x_1 - \frac{\partial \Phi}{\partial y_1} y_1 - \dots \\ T_{21} &= \quad - \frac{\partial \Phi}{\partial x_1} x_2 - \frac{\partial \Phi}{\partial y_1} y_2 - \dots \ . \end{aligned}\right\} \tag{12}$$

If we denote the state variables used in the calculation with H by $\xi_1, \xi_2, \xi_3, \xi_4$; $\eta_1, \eta_2, \eta_3, \eta_4$; $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ etc., then the following differential equation follows from the definition of Φ (*Theory of Matter*, Part III, formulas (98) and (93) on page 32 and 30):

$$\frac{\partial}{\partial \omega_{\mu\nu}} H(\xi_1, \xi_2, \dots, \omega_{11}, \omega_{12}, \dots) = \frac{\partial}{\partial \omega_{\mu\nu}} \Phi(x_1, x_2, \dots, \omega_{11}, \omega_{12}, \dots).$$

Accordingly, by differentiating (12) and using (10) we obtain:

$$\begin{aligned} \frac{1}{\kappa} \frac{\partial T_{11}}{\partial \omega_{21}} &= - T_{21} + \frac{\partial T_{21}}{\partial x_1} x_1 + \frac{\partial T_{21}}{\partial y_1} y_1 + \dots \\ \frac{1}{\kappa} \frac{\partial T_{21}}{\partial \omega_{11}} &= \qquad \frac{\partial T_{11}}{\partial x_1} x_2 + \frac{\partial T_{11}}{\partial y_1} y_2 + \dots \ . \end{aligned}$$

But it is easy to see that

$$\frac{\partial T_{11}}{\partial x_1} x_2 + \frac{\partial T_{11}}{\partial y_1} y_2 + \dots = \frac{\partial T_{21}}{\partial x_1} x_1 + \frac{\partial T_{21}}{\partial y_1} y_1 + \dots \ .$$

Therefore eq. (10) also leads to:

$$\frac{\partial T_{11}}{\partial \omega_{21}} + \kappa T_{21} = \frac{\partial T_{21}}{\partial \omega_{11}} \ . \tag{13}$$

Equation (11) and eq. (13) can be simultaneously satisfied only if $T_{21} = 0$. Exactly the same proof applies for any arbitrary $T_{\mu\nu}$ $(\mu \neq \nu)$. So the principle of identity of the two masses leads to the conclusion that all off-diagonal components of the energy tensor are equal to zero. But that would only be possible if the energy tensor were in reality a scalar. This proves the impossibility of the identity of the two masses.

The principle of identity of gravitational and inertial mass is impossible also in a theory in which the gravitational potential and the density of gravitational mass are four-dimensional tensors.

Indeed a glance at the formulas (15), (18) as well as (19) on p. 16 and 17 of Einstein's treatise shows that also in Einstein's theory the tensor for the denseness of gravitational mass found in formulas (15) and (18) is quite different from the energy tensor given by (19). It is therefore an error when Mr. Einstein speaks in the cited

treatise of a "*physical unity of essence of gravitational and inertial mass*" in his theory, and of the validity of the "*equivalence hypothesis*," according to which "*the identity of gravitational and inertial mass is satisfied exactly*."

Nevertheless, there still remains the possibility that the gravitational and inertial mass in large bodies consisting of molecules can be made strictly proportional to each other by means of a series of auxiliary assumptions leading to a compensation of the deviations in the separate elementary particles when integrating over the whole volume. Indeed it seems to me that Mr. Einstein had only this remaining possibility in mind in his Vienna lecture,[6] when he postulated the equality of inertial and gravitational mass for "closed systems" (§ 2, postulate 2).

By itself it is probably rather immaterial whether one can, in such a somewhat artificial way, get the two masses to be mathematically exactly equal in closed systems, or whether they are only approximately equal, once one has abandoned the identity of the two masses in principle; whereby then, after all, the thoughts about general relativity of motion, of which Mr. Einstein spoke in such detail in his lecture, are abandoned as well.

Since Mr. Einstein puts so much weight on the validity of the theorem of the equality of the two masses in his theory, at least for closed systems, we are forced to go into the details of this theorem when discussing his theory.

4. SPECIAL ASSUMPTIONS OF EINSTEIN'S THEORY

In the cited paper Mr. Einstein uses the notation $g_{\mu\nu}$ for the components of the gravitational potential. The quantities that we denoted in eq. (1) by $\omega_{\mu\nu}$ differ from the $g_{\mu\nu}$ only by a constant factor:

$$g_{\mu\nu} = -2\kappa \cdot \omega_{\mu\nu}. \tag{14}$$

In a region infinitely distant from all matter, that is, in an ideal vacuum, the tensor $(g_{\mu\nu})$ is supposed to degenerate into the scalar -1, so that its several components take the following values: |

$$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix} \qquad [119]$$

Following Minkowski I always put the speed of light in an ideal vacuum equal to 1. In order to arrive at the special theory of Einstein one must make the following assumptions:

1. The Hamiltonian function H can be split into two terms, $H = H_e + H_g$, both of which depend on the components of the gravitational potential, that is, on the

6 *Physik. Zeitschr.* 14, 1249, 1913 [in this volume].

quantities $g_{\mu\nu}$; but additionally, the second one, H_g, contains only the components of the vector of tensors of the gravitational field $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$, and the first one, H_e, contains only the remaining state variables, e.g., the electromagnetic field quantities etc.

2. The dependence of the quantity H_e on the $g_{\mu\nu}$ shall take the form of the following expression:

$$\left.\begin{aligned} H_e &= \sqrt{g_{11}X_{11} + g_{22}X_{22} + \cdot\cdot + 2g_{12}X_{12} + \cdot\cdot} \\ H_e &= \sqrt{\sum_{\mu}\sum_{\nu} g_{\mu\nu}X_{\mu\nu}}\,, \end{aligned}\right\} \tag{15}$$

where the $X_{\mu\nu}$ no longer contain the $g_{\mu\nu}$, but only the other state variables of the aether. The $X_{\mu\nu}$ form the components of a four-dimensional tensor, H_e is then a four-dimensional scalar.

3. In the case that the material particle we consider is at rest, the tensor $(X_{\mu\nu})$ reduces to a tensor all of whose components are zero except X_{44}. Following Einstein we shall denote this value X_{44}^0 by $-\rho^2$:

$$X_{44}^0 = -\rho^2.$$

So the quantity ρ is a four-dimensional scalar. If the particle moves with velocity $\mathfrak{q}$, then we define in the usual way the following quantity $\mathfrak{B}$ as the velocity four vector:

$$\mathfrak{B}_1 = \frac{\mathfrak{q}_x}{\sqrt{1-\mathfrak{q}^2}}\,, \qquad \mathfrak{B}_2 = \frac{\mathfrak{q}_y}{\sqrt{1-\mathfrak{q}^2}}\,, \qquad \mathfrak{B}_3 = \frac{\mathfrak{q}_z}{\sqrt{1-\mathfrak{q}^2}}\,,$$

$$\mathfrak{B}_4 = \frac{i}{\sqrt{1-\mathfrak{q}^2}}\,, \qquad \mathfrak{B}_1{}^2 + \mathfrak{B}_2{}^2 + \mathfrak{B}_3{}^2 + \mathfrak{B}_4{}^2 = -1\,.$$

It then follows directly from the principle of relativity that the components of the tensor $(X_{\mu\nu})$ take the following values in the case that the particle we consider is in motion:

$$X_{\mu\nu} = +\rho^2\mathfrak{B}_\mu\mathfrak{B}_\nu. \tag{16}$$

Accordingly, the first term of the Hamiltonian function has the value:

$$\left.\begin{aligned} H_e &= \rho\sqrt{(g_{11}\mathfrak{B}_1{}^2 + g_{22}\mathfrak{B}_2{}^2 + \cdot\cdot + 2g_{12}\mathfrak{B}_1\mathfrak{B}_2 + \cdot\cdot)} \\ H_e &= \rho\sqrt{\sum_{\mu}\sum_{\nu} g_{\mu\nu}\mathfrak{B}_\mu\mathfrak{B}_\nu}\,, \end{aligned}\right\} \tag{17}$$

H_e and ρ are four-dimensional scalars as is the square root term in eq. (17),[3] so that we have:

$$\sum_{\mu}\sum_{\nu} g_{\mu\nu}\mathfrak{P}_{\mu}\mathfrak{P}_{\nu} = -g_{44}^{0}. \tag{18}$$

4. The second term of the Hamiltonian function H_g is a homogeneous function of second degree in the components of the vector of tensors $(\mathfrak{g}_{\mu\nu}, iu_{\mu\nu})$. If we write for simplicity the letters: $\mathfrak{g}_{\mu\nu 1}, \mathfrak{g}_{\mu\nu 2}, \mathfrak{g}_{\mu\nu 3}, \mathfrak{g}_{\mu\nu 4}$ for the quantities $\mathfrak{g}_{\mu\nu x}, \mathfrak{g}_{\mu\nu y}, \mathfrak{g}_{\mu\nu z}, iu_{\mu\nu}$, this means that H_g is an expression of the following form

$$H_g = \frac{1}{2}\cdot \sum_{\kappa,\,\mu,\,\lambda,\,\nu,\,\alpha,\,\beta} G_{\kappa\mu\lambda\nu\alpha\beta}\mathfrak{g}_{\kappa\lambda\alpha}\mathfrak{g}_{\mu\nu\beta}, \tag{19}$$

where each of the six indices is to be summed over separately from 1 to 4. The coefficients $G_{\kappa\mu\lambda\nu\alpha\beta}$ are functions of the gravitational potential, that is of the quantities $g_{\mu\nu}$, about which we have to make certain stipulations in order to come to Einstein's theory.

5. The tensor $(\gamma_{\mu\nu})$ defined by the following equations:

$$\left.\begin{aligned} g_{1\nu}\gamma_{1\nu} + g_{2\nu}\gamma_{2\nu} + g_{3\nu}\gamma_{3\nu} + g_{4\nu}\gamma_{4\nu} &= 1 \\ g_{1\mu}\gamma_{1\nu} + g_{2\mu}\gamma_{2\nu} + g_{3\mu}\gamma_{3\nu} + g_{4\mu}\gamma_{4\nu} &= 0, \qquad \mu \neq \nu \end{aligned}\right\} \tag{20}$$

is called the *inverse tensor* [*reziproker Tensor*] of $(g_{\mu\nu})$. Furthermore, let:

$$g = \begin{vmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{vmatrix}. \tag{21}$$

Then $\gamma_{\mu\nu}$ can be calculated as the cofactor [*adjungierte Unterdeterminante*] of $g_{\mu\nu}$ divided by g.[7] Let us now put:

$$G_{\kappa\mu\lambda\nu\alpha\beta} = 2\kappa\sqrt{g}\gamma_{\kappa\mu}\gamma_{\lambda\nu}\gamma_{\alpha\beta}. \tag{22}$$

5. THE FUNDAMENTAL EQUATIONS OF EINSTEIN'S THEORY

Before we substitute the expression, that results according to the above stipulations for H, into the general eqs. (1) and (2) in accordance with (5) and (6), we want to put down a few simple formulas which result directly from the eqs. (20) defining $\gamma_{\mu\nu}$, and which are very convenient for the following calculations. Differentiation of (20) with respect to any coordinate x_β yields:[4]

---

7 I take the sign of g as positive. Einstein has some signs different from me, because he puts $x_4 = t$ whereas I put $x_4 = it$.

$$g_{1\lambda}\frac{\partial \gamma_{1\nu}}{\partial x_\beta} + g_{2\lambda}\frac{\partial \gamma_{2\nu}}{\partial x_\beta} + g_{3\lambda}\frac{\partial \gamma_{3\nu}}{\partial x_\beta} + g_{4\lambda}\frac{\partial \gamma_{4\nu}}{\partial x_\beta} = -\left(\gamma_{1\nu}\frac{\partial g_{1\lambda}}{\partial x_\beta} + \gamma_{2\nu}\frac{\partial g_{2\lambda}}{\partial x_\beta} + \gamma_{3\nu}\frac{\partial g_{3\lambda}}{\partial x_\beta} + \gamma_{4\nu}\frac{\partial g_{4\lambda}}{\partial x_\beta}\right) \tag{23}$$

for $\lambda = \nu$ as well as for $\lambda \neq \nu$. |

[120] Now we write down the 4 eqs. (23) for a particular value ν, putting sequentially $\lambda = 1, 2, 3, 4$, we multiply each equation by $\gamma_{\mu\lambda}$, where μ is another constant number which may be different from or equal to ν, and then we add, with the result:

$$\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta} = -\sum_\kappa \sum_\lambda \gamma_{\mu\lambda}\gamma_{\kappa\nu}\frac{\partial g_{\kappa\lambda}}{\partial x_\beta}. \tag{24}$$

By differentiating (20) with respect to any g_{mn}, treating g_{mn} and g_{nm} as different variables, we obtain:[5]

$$\left.\begin{aligned} \sum_{\kappa=1}^{4} g_{\kappa\lambda}\frac{\partial \gamma_{\kappa\nu}}{\partial g_{mn}} &= 0, \quad \lambda \neq n \\ \sum_{\kappa=1}^{4} g_{\kappa n}\frac{\partial \gamma_{\kappa\nu}}{\partial g_{mn}} + \gamma_{m\nu} &= 0. \end{aligned}\right\} \tag{25}$$

For fixed ν, we multiply the four equations for $\lambda = 1, 2, 3, 4$ by $\gamma_{\mu\lambda}$ and add to obtain:

$$\frac{\partial \gamma_{\mu\nu}}{\partial g_{mn}} + \gamma_{m\nu}\gamma_{\mu n} = 0. \tag{26}$$

Finally let us differentiate the determinant g with respect to g_{mn}, where again g_{mn} and g_{nm} are treated as different variables:

$$\frac{\partial g}{\partial g_{mn}} = g \cdot \gamma_{mn}. \tag{27}$$

From (27) it follows that:

$$\frac{\partial \sqrt{g}}{\partial g_{mn}} = \frac{1}{2}\sqrt{g}\gamma_{mn}. \tag{28}$$

After these preparations we now calculate first the vector of tensors $(\mathfrak{k}_{mn}, iw_{mn})$ and then the tensor of gravitational mass h_{mn}.

I will call the four quantities $\mathfrak{k}_{mnx}, \mathfrak{k}_{mny}, \mathfrak{k}_{mnz}, iw_{mn}$ for simplicity $\mathfrak{k}_{mn1}, \mathfrak{k}_{mn2}, \mathfrak{k}_{mn3}, \mathfrak{k}_{mn4}$. From (5) and (19) one then has:

$$\mathfrak{k}_{mn\alpha} = \frac{\partial H}{\partial \mathfrak{g}_{mn\alpha}} = \sum_{\mu,\nu,\beta} G_{m\mu n\nu\alpha\beta}\mathfrak{g}_{\mu\nu\beta}.$$

Here again the $\mathfrak{f}_{mn}$ and $\mathfrak{f}_{nm}$ as well as the $\mathfrak{g}_{mn}$ and $\mathfrak{g}_{nm}$ are regarded as different variables, but after the differentiation we set $G_{m\mu n\nu\alpha\beta} = G_{\mu m\nu n\beta\alpha}$. If we write, from (1) and (14),

$$\mathfrak{g}_{mn\beta} = -\frac{1}{2\kappa}\frac{\partial g_{mn}}{\partial x_\beta},$$

then we find, using formulas (22) and (24),

$$\mathfrak{f}_{mn\alpha} = \sum_{\beta=1}^{4} \sqrt{g}\gamma_{\alpha\beta}\frac{\partial \gamma_{mn}}{\partial x_\beta}. \tag{29}$$

This is the same vector of tensors appearing in Einstein's paper in the differential equation for the gravitational field, (15) and (18) on p. 16 and 17. It is easily seen that H_g can be written in the following form:

$$H_g = \frac{1}{2}\sum_{m,n,\alpha} \mathfrak{f}_{mn\alpha}\mathfrak{g}_{mn\alpha}. \tag{30}$$

Now we turn to the calculation of h_{mn}. From (6) and (14) we have:

$$h_{mn} = -\frac{1}{\kappa}\frac{\partial H}{\partial \omega_{mn}} = +2\frac{\partial H}{\partial g_{mn}}. \tag{31}$$

Since $H = H_e + H_g$, h_{mn} also splits into a sum: $h_{mn} = h^e_{mn} + h^g_{mn}$. From (16) and (17) we have

$$\left.\begin{aligned} h^e_{mn} &= \frac{\rho \mathfrak{B}_m \mathfrak{B}_n}{\sqrt{\sum_{\mu,\nu} g_{\mu\nu}\mathfrak{B}_\mu \mathfrak{B}_\nu}} \\ h^e_{mn} &= \frac{\rho}{\sqrt{-g^0_{44}}}\mathfrak{B}_m \mathfrak{B}_n. \end{aligned}\right\} \tag{32}$$

So h^e_{mn} has the form $h^e_{mn} = h\mathfrak{B}_m\mathfrak{B}_n$, where $h = \rho/\sqrt{-g^0_{44}}$ is a four-dimensional scalar. To compare (32) with Einstein's expression for h^e_{mn}, I put (Einstein, eq. (1″), p. 6):

$$\sqrt{\sum_{\mu,\nu} g_{\mu\nu}dx_\mu dx_\nu} = ds;$$

if furthermore dV is the volume element at the spacetime point under consideration, then the rest volume dV' of the mass particle occupying dV is

$$dV' = \frac{dV}{\sqrt{1-\mathfrak{q}^2}},$$

and I define:

$$dm = \rho dV' = \rho \frac{dV}{\sqrt{1-\mathfrak{q}^2}}.$$

Further, following Einstein (p. 10), I put

$$dV_0 = \sqrt{g} \cdot \frac{dV}{\frac{ds}{dt}},$$

and

$$\rho_0 = \frac{dm}{dV_0} = \frac{\rho}{\sqrt{g}\sqrt{1-\mathfrak{q}^2}} \frac{ds}{dt}.$$

We then have:

$$h^e_{mn} = \rho_0 \sqrt{g} \frac{dx_m}{ds} \frac{dx_n}{ds},$$

or, as in Einstein (p. 10),[8] putting

$$\Theta_{mn} = -\rho_0 \frac{dx_m}{ds} \frac{dx_n}{ds},$$

[121] | we have:

$$h^e_{mn} = -\sqrt{g} \cdot \Theta_{mn} \tag{33}$$

in agreement with Einstein's eqs. (15) and (18).

Now I calculate the second term of h_{mn}. From (19) and (31) we have

$$h^g_{mn} = \sum_{\kappa,\,\mu,\,\lambda,\,\nu,\,\alpha,\,\beta} \frac{\partial G_{\kappa\mu\lambda\nu\alpha\beta}}{\partial g_{mn}} \mathfrak{g}_{\kappa\lambda\alpha} \mathfrak{g}_{\mu\nu\beta},$$

where

$$G_{\kappa\mu\lambda\nu\alpha\beta} = 2\kappa\sqrt{g}\gamma_{\kappa\mu}\gamma_{\lambda\nu}\gamma_{\alpha\beta}.$$

When differentiating, g_{mn} and g_{nm} are to be treated as different. The formulas (26) and (28) yield:

$$\frac{\partial G_{\kappa\mu\lambda\nu\alpha\beta}}{\partial g_{mn}} = \kappa\sqrt{g}\gamma_{mn}\gamma_{\kappa\mu}\gamma_{\lambda\nu}\gamma_{\alpha\beta} - 2\kappa\sqrt{g}\gamma_{\alpha n}\gamma_{m\beta}\gamma_{\kappa\mu}\gamma_{\lambda\nu} -$$
$$-2\kappa\sqrt{g}(\gamma_{\alpha\beta}\gamma_{\lambda n}\gamma_{m\nu}\gamma_{\kappa\mu} + \gamma_{\alpha\beta}\gamma_{\kappa n}\gamma_{m\mu}\gamma_{\lambda\nu}).$$

8 I choose the negative sign to make Θ_{44} positive.

If one further puts

$$\mathfrak{g}_{\kappa\lambda\alpha} = -\frac{1}{2\kappa}\frac{\partial g_{\kappa\lambda}}{\partial x_\alpha}, \qquad \mathfrak{g}_{\mu\nu\beta} = -\frac{1}{2\kappa}\frac{\partial g_{\mu\nu}}{\partial x_\beta},$$

and notes that $g_{\mu\nu} = g_{\nu\mu}$, $\gamma_{\mu\nu} = \gamma_{\nu\mu}$, then, with the aid of formula (24), one finds:

$$\begin{aligned} h^g_{mn} = &-\frac{\sqrt{g}}{4\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{mn}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} + \frac{\sqrt{g}}{2\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha m}\gamma_{\beta n}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} \\ &+ \frac{\sqrt{g}}{\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{m\nu}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}, \end{aligned} \tag{34}$$

or from formula (23)

$$\begin{aligned} h^g_{mn} = &-\frac{\sqrt{g}}{4\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{mn}\gamma_{\alpha\beta}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} + \frac{\sqrt{g}}{2\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha m}\gamma_{\beta n}\frac{\partial g_{\mu\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta} \\ &- \frac{\sqrt{g}}{\kappa}\sum_{\mu,\nu,\alpha,\beta}g_{\mu\nu}\gamma_{\alpha\beta}\frac{\partial\gamma_{m\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}. \end{aligned}$$

This last form is used by Einstein; he collects together the first two terms calling their sum $-\sqrt{g}\cdot\vartheta_{mn}$ (Einstein, p. 15, formula (13)), and writes:

$$h^g_{mn} = -\sqrt{g}\left(\vartheta_{mn} + \frac{1}{\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha\beta}g_{\mu\nu}\frac{\partial\gamma_{m\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}\right). \tag{35}$$

Written in Einstein's notation, the density of the gravitational mass is then, taken together:

$$h_{mn} = -\sqrt{g}\left(\Theta_{mn} + \vartheta_{mn} + \frac{1}{\kappa}\sum_{\mu,\nu,\alpha,\beta}\gamma_{\alpha\beta}g_{\mu\nu}\frac{\partial\gamma_{m\nu}}{\partial x_\alpha}\frac{\partial\gamma_{\mu n}}{\partial x_\beta}\right). \tag{36}$$

Taking into account formula (29),

$$\mathfrak{k}_{mn\alpha} = \sum_\beta\sqrt{g}\gamma_{\alpha\beta}\frac{\partial\gamma_{mn}}{\partial x_\beta},$$

one sees immediately that Einstein's fundamental eq. (15) and (18) on p. 16 and p. 17 is nothing other than our eq. (2):

$$\sum_\alpha\frac{\partial\mathfrak{k}_{mn\alpha}}{\partial x_\alpha} = -\kappa h_{mn}.$$

Einstein's theory of gravitation is a special case of the general theory of gravitation with a tensor potential described in section 2. Thus it fits perfectly into the framework of the ordinary theory of relativity.

6. THE ENERGY TENSOR

In order to set up the energy tensor, Einstein adds another auxiliary assumption to those enumerated in Section 4, which will become particularly important in the course of our investigation. Let us define a new four-vector $\mathfrak{W}$ as follows:

$$\left.\begin{aligned} \mathfrak{W}_\nu &= \frac{\partial H_e}{\partial \mathrm{V}_\nu} = \frac{\rho \sum_\mu g_{\mu\nu}\mathfrak{V}_\mu}{\sqrt{\sum_\mu \sum_\nu g_{\mu\nu}\mathfrak{V}_\mu \mathfrak{V}_\nu}}, \\ \mathfrak{W}_\nu &= \frac{\rho}{\sqrt{-g^0_{44}}} \sum_{\mu=1}^{4} g_{\mu\nu}\mathfrak{V}_\mu \,. \end{aligned}\right\} \tag{37}$$

The new auxiliary assumption of Einstein is:

$$\begin{aligned} &\frac{\partial}{\partial x}(\mathfrak{W}_\mu \cdot \mathfrak{V}_1) + \frac{\partial}{\partial y}(\mathfrak{W}_\mu \cdot \mathfrak{V}_2) + \frac{\partial}{\partial z}(\mathfrak{W}_\mu \cdot \mathfrak{V}_3) - i\frac{\partial}{\partial t}(\mathfrak{W}_\mu \cdot \mathfrak{V}_4) \\ &\qquad = -\kappa \sum_{m,n} \mathfrak{g}_{mn\mu} h^e_{mn} = \sum_{m,n} \frac{\partial H_e}{\partial g_{mn}} \frac{\partial g_{mn}}{\partial x_\mu}. \end{aligned} \tag{38}$$

Einstein calls this auxiliary assumption the law of energy-momentum conservation. It is found as eq. (10) on p. 10 in his treatise, and there takes the form

$$\sum_{\mu,\nu} \frac{\partial}{\partial x_\nu}(\sqrt{g}\, g_{\sigma\mu}\Theta_{\mu\nu}) - \frac{1}{2}\sum \sqrt{g}\frac{\partial g_{\mu\nu}}{\partial x_\sigma}\Theta_{\mu\nu} = 0. \quad |$$

[122] Noting that as a consequence of our eqs. (32) and (33):

$$\mathfrak{W}_\sigma \mathfrak{V}_\nu = -\sqrt{g}\sum_\mu g_{\sigma\mu}\Theta_{\mu\nu},$$

and that according to (31) and (33)

$$\frac{\partial H_e}{\partial g_{\mu\nu}} = -\frac{1}{2}\sqrt{g}\Theta_{\mu\nu},$$

one sees immediately that this equation is identical with (38).

Assuming (38) to be correct, one easily finds that the components of a tensor defined as follows:

$$\left.\begin{aligned} T_{\alpha\alpha} &= \mathfrak{W}_\alpha \mathfrak{V}_\alpha + H_g - \sum_\mu \sum_\nu \mathfrak{g}_{\mu\nu\alpha}\mathfrak{f}_{\mu\nu\alpha}, \\ T_{\alpha\beta} &= \mathfrak{W}_\alpha \mathfrak{V}_\beta - \sum_\mu \sum_\nu \mathfrak{g}_{\mu\nu\alpha}\mathfrak{f}_{\mu\nu\beta} \end{aligned}\right\} \tag{39}$$

satisfy the differential equations

$$\frac{\partial T_{\alpha 1}}{\partial x} + \frac{\partial T_{\alpha 2}}{\partial y} + \frac{\partial T_{\alpha 3}}{\partial z} - i\frac{\partial T_{\alpha 4}}{\partial t} = 0; \tag{40}$$

for the proof one uses eqs. (38), (1), (2), (5), (6). So the tensor (39) is the energy tensor, the eqs. (40) are the energy-momentum equations. A bit of calculation shows that the eqs. (40) are identical with the eqs. (19) on p. 17 of Einstein:

$$\sum_{\mu,\nu} \frac{\partial}{\partial x_\nu}(\sqrt{g}\, g_{\sigma\mu}(\Theta_{\mu\nu} + \vartheta_{\mu\nu})) = 0, \qquad (\sigma = 1, 2, 3, 4.)$$

In a gravitation-free space, where $g_{\nu\nu} = -1$, $g_{\mu\nu} = 0$, $\mathfrak{g}_{\mu\nu\alpha} = 0$, the components of the energy tensor take on the following values:

$$T_{\alpha\alpha}{}' = -\rho \mathfrak{B}_\alpha{}^2, \qquad T_{\alpha\beta}{}' = -\rho \mathfrak{B}_\alpha \mathfrak{B}_\beta.$$

Thus, according to eqs. (16), the tensor $(X_{\mu\nu})$ introduced on p. 17 differs from the energy tensor $(T_{\mu\nu}{}')$ in gravitation free space only by the factor $-\rho$,

$$X_{\mu\nu} = -\rho T_{\mu\nu}{}'.$$

By decomposing the components of the energy tensor into two terms in the same way as $H = H_e + H_g$,

$$T_{\alpha\beta} = T^e_{\alpha\beta} + T^g_{\alpha\beta},$$

where

$$T^e_{\alpha\beta} = \mathfrak{W}_\alpha \mathfrak{B}_\beta$$

and

$$T^g_{\alpha\beta} = -\sum_{\mu,\nu} \mathfrak{g}_{\mu\nu\alpha} \mathfrak{f}_{\mu\nu\beta},$$

or

$$T^g_{\alpha\alpha} = H_g - \sum_{\mu,\nu} \mathfrak{g}_{\mu\nu\alpha} \mathfrak{f}_{\mu\nu\alpha},$$

one notes directly that (16) and (37) imply the theorem:

$$H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}. \tag{41}$$

The auxiliary assumptions (16) and (38) of Einstein's theory imply that that part of the Hamiltonian function H_e which does not contain the field strength of gravity becomes identical to the sum of the diagonal terms of the part of the energy tensor that is devoid of gravitational field strength.

[169] 7. THEOREM OF THE RELATIVITY OF THE GRAVITATIONAL POTENTIAL

The assumptions of Einstein's theory are perhaps in part of rather secondary importance, but they are all made according to one principle, namely that the resulting equations admit other linear transformations in addition to the Lorentz transformations. I believe one may characterize the two propositions:

1. *that the gravitational potential is a four-dimensional tensor,*
2. *that the general equations of the aether dynamics admit other linear transformations in addition to those of Lorentz*,

as the two essential or main assumptions of Einstein's theory, compared to which the other assumptions play a subsidiary role as auxiliary assumptions.

To grasp the true nature of the second proposition, which Einstein regards as a generalization of the principle of relativity, it will be necessary to go into it rather precisely, although in doing so repetition of several calculations of Messrs. A. Einstein and M. Grossmann will be unavoidable.

Let us imagine a material system located in empty space, far distant from all other matter, that is at a place where the gravitational potential has the scalar value -1, and that we have complete knowledge of the processes and their laws in this system. Further, we imagine this same material system transported into the vicinity of a very large body, the Earth for example, where the gravitational potential is no longer equal to -1, but instead is represented by a tensor. Because the gravitational potential enters into the function H and thereby also into the equations of the aether dynamics, presumably all processes in the material system are influenced by the mere presence of a gravitational potential that differs from -1. Now the question is, what is the nature of this influence of the gravitational potential. This question is one that I, too, have already asked in my theory of the scalar gravitational potential, and I have given an answer for the case of that theory (loc. cit. III, p. 61ff).

At the second location the field strength of gravitation shall also be so small that the changes of the gravitational potential do not reach any appreciable values at the boundaries of a region containing the material system and extending infinitely far in comparison with that system; that is, the gravitational potential may be considered constant on the boundary. Let us denote by $(g^{1}_{\mu\nu})$ this constant value of the gravitational potential at infinite distance from the material system being considered. By the following equations we will then define 16 transformation coefficients $a_{\mu\nu}$, which
[170] together form a four dimensional and in general asymmetric | matrix $(a_{\mu\nu} \neq a_{\nu\mu})$:

$$a_{\mu 1}a_{\nu 1} + a_{\mu 2}a_{\nu 2} + a_{\mu 3}a_{\nu 3} + a_{\mu 4}a_{\nu 4} = -g^{1}_{\mu\nu}. \tag{42}$$

Since these are only 10 equations, 6 of the coefficients $a_{\mu\nu}$ can of course be chosen arbitrarily. I will denote the inverse matrix of the matrix $(a_{\mu\nu})$ by $(\alpha_{\mu\nu})$, so the elements $\alpha_{\mu\nu}$ are defined as follows:

$$\left.\begin{aligned} &a_{1\mu}\alpha_{1\mu} + a_{2\mu}\alpha_{2\mu} + a_{3\mu}\alpha_{3\mu} + a_{4\mu}\alpha_{4\mu} = 1, \\ &a_{1\mu}\alpha_{1\nu} + a_{2\mu}\alpha_{2\nu} + a_{3\mu}\alpha_{3\nu} + a_{4\mu}\alpha_{4\nu} = 0, \qquad \mu \neq \nu \end{aligned}\right\} \tag{43}$$

Now I introduce in place of the components of the gravitational potential $g_{\mu\nu}$ in the interior and in the closer vicinity of said material system the following linear functions of the $g_{\mu\nu}$, which I will call $g_{\mu\nu}'$:

$$g_{\mu\nu}' = \sum_{\kappa,\lambda} \alpha_{\kappa\mu}\alpha_{\lambda\nu}g_{\kappa\lambda}. \tag{44}$$

The ten quantities $g_{\mu\nu}'$ taken together again form a four-dimensional tensor, which originated from the tensor $(g_{\mu\nu})$ by deformation and rotation, as it were; at infinity, where $(g_{\mu\nu})$ reaches $(g^{1}_{\mu\nu})$, $(g_{\mu\nu}')$ becomes, by formulas (42) and (43), the scalar -1. Further I calculate $(\gamma_{\mu\nu}')$, the inverse tensor to $(g_{\mu\nu}')$, defined by:

$$\left.\begin{aligned} &g_{\mu 1}'\gamma_{\mu 1}' + g_{\mu 2}'\gamma_{\mu 2}' + g_{\mu 3}'\gamma_{\mu 3}' + g_{\mu 4}'\gamma_{\mu 4}' = 1, \\ &g_{\mu 1}'\gamma_{\nu 1}' + g_{\mu 2}'\gamma_{\nu 2}' + g_{\mu 3}'\gamma_{\nu 3}' + g_{\mu 4}'\gamma_{\nu 4}' = 0, \quad \mu \neq \nu. \end{aligned}\right\}$$

It is easily seen that the $\gamma_{\mu\nu}'$ can be represented as linear functions of the components $\gamma_{\mu\nu}$ of the inverse tensor to $g_{\mu\nu}$:

$$\gamma_{\mu\nu}' = \sum_{\kappa,\lambda} a_{\kappa\mu}a_{\lambda\nu}\gamma_{\kappa\lambda}. \tag{45}$$

Using formulas (43) one can easily verify the equations of definition of the $\gamma_{\mu\nu}'$. Conversely, if one wants to calculate the $g_{\mu\nu}$ from the $g_{\mu\nu}'$ and the $\gamma_{\mu\nu}$ from the $\gamma_{\mu\nu}'$, one has:

$$g_{\mu\nu} = \sum_{\kappa,\lambda} a_{\mu\kappa}a_{\nu\lambda}g_{\kappa\lambda}', \qquad \gamma_{\mu\nu} = \sum_{\kappa,\lambda} \alpha_{\mu\kappa}\alpha_{\nu\lambda}\gamma_{\kappa\lambda}'. \tag{46}$$

In place of the rectangular coordinate system (x_1, x_2, x_3, x_4) we introduce further an oblique-angled one (x_1', x_2', x_3', x_4'), which moreover has different units of lengths on the different coordinate axes, by making the following substitutions:

$$x_\nu = \alpha_{\nu 1}x_1' + \alpha_{\nu 2}x_2' + \alpha_{\nu 3}x_3' + \alpha_{\nu 4}x_4'. \tag{47}$$

We then have:

$$x_\nu' = a_{1\nu}x_1 + a_{2\nu}x_2 + a_{3\nu}x_3 + a_{4\nu}x_4. \tag{48}$$

Further we denote by g the determinant of the $g_{\mu\nu}$, as in eq. (21) above, similarly by g_1 and g' the determinants of the $g^{1}_{\mu\nu}$ and the $g_{\mu\nu}'$. Then it follows directly from (42) and (46) that:

$$g = g_1 g'. \tag{49}$$

We further define:

$$\mathfrak{g}_{\mu\nu\alpha}' = -\frac{1}{2\kappa}\frac{\partial g_{\mu\nu}'}{\partial x_\alpha'}, \qquad \mathfrak{f}_{\mu\nu\alpha}' = \sqrt{g'}\sum_\beta \gamma_{\alpha\beta}'\frac{\partial \gamma_{\mu\nu}'}{\partial x_\beta'}, \tag{50}$$

and by analogy to (30):

$$H_g' = \frac{1}{2}\sum_{\mu,\nu,\alpha}\mathfrak{g}_{\mu\nu\alpha}'\mathfrak{f}_{\mu\nu\alpha}'. \tag{51}$$

By a simple calculation it can then be shown that:

$$H_g' = \frac{1}{\sqrt{g_1}}H_g. \tag{52}$$

If we regard H_g' as a function of the transformed quantities $g_{\mu\nu}'$ and $\mathfrak{g}_{\mu\nu\alpha}'$, then (51) implies:

$$\mathfrak{f}_{\mu\nu\alpha}' = \frac{\partial H_g'}{\partial \mathfrak{g}_{\mu\nu\alpha}'}. \tag{53}$$

Next we introduce a new velocity vector $\mathfrak{V}'$, obtained from $\mathfrak{V}$ by the following transformation equations:

$$\frac{\mathfrak{V}_\mu}{s} = \alpha_{\mu1}\mathfrak{V}_1' + \alpha_{\mu2}\mathfrak{V}_2' + \alpha_{\mu3}\mathfrak{V}_3' + \alpha_{\mu4}\mathfrak{V}_4', \tag{54}$$

or:

$$s\mathfrak{V}_\mu' = a_{1\mu}\mathfrak{V}_1 + a_{2\mu}\mathfrak{V}_2 + a_{3\mu}\mathfrak{V}_3 + a_{4\mu}\mathfrak{V}_4, \tag{55}$$

where s is to denote the following quantity:

$$s^2 = \sum_{\mu,\nu} g^1_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu = \frac{1}{\sum_{\mu,\nu}\gamma^1_{\mu\nu}\mathfrak{V}_\mu'\mathfrak{V}_\nu'}. \tag{56}$$

By squaring and adding the eqs. (55) one finds, taking note of (42):

$$\mathfrak{V}_1'^2 + \mathfrak{V}_2'^2 + \mathfrak{V}_3'^2 + \mathfrak{V}_4'^2 = -1. \tag{57}$$

It is easily seen that

$$\sum_{\mu,\nu} g_{\mu\nu}\mathfrak{V}_\mu\mathfrak{V}_\nu = s^2\sum_{\mu,\nu} g_{\mu\nu}'\mathfrak{V}_\mu'\mathfrak{V}_\nu',$$

hence by (16):

$$H_e = \rho\, s\sqrt{\sum_{\mu,\nu} g_{\mu\nu}'\mathfrak{V}_\mu'\mathfrak{V}_\nu'}.$$

Let us next define:

$$\rho' = \frac{\rho s}{\sqrt{g_1}} = \frac{\rho}{\sqrt{g_1}}\sqrt{\sum_{\mu,\nu} g^1_{\mu\nu}\mathfrak{B}_\mu\mathfrak{B}_\nu}\,, \tag{58}$$

$$H_e' = \rho'\sqrt{\sum_{\mu,\nu} g_{\mu\nu}'\mathfrak{B}_\mu'\mathfrak{B}_\nu'}\,, \tag{59}$$ [171]

then we have:

$$H_e' = \frac{1}{\sqrt{g_1}}H_e, \tag{60}$$

and from (52):

$$H' = H_e' + H_g' = \frac{1}{\sqrt{g_1}}H. \tag{61}$$

Now we define, by analogy to (37):

$$\mathfrak{W}_\nu' = \frac{\partial H_e'}{\partial \mathfrak{B}_\nu'} = \frac{\rho'\sum_\mu g_{\mu\nu}'\mathfrak{B}_\mu'}{\sqrt{\sum_{\mu,\nu} g_{\mu\nu}'\mathfrak{B}_\mu'\mathfrak{B}_\nu'}}, \tag{62}$$

then a simple calculation shows that:

$$\mathfrak{W}_\nu' = \frac{s}{\sqrt{g_1}}(\alpha_{1\nu}\mathfrak{W}_1 + \alpha_{2\nu}\mathfrak{W}_2 + \alpha_{3\nu}\mathfrak{W}_3 + \alpha_{4\nu}\mathfrak{W}_4). \tag{63}$$

After these preliminaries we now quickly come to the conclusion. From the definition (53) of the $\mathfrak{f}_{\mu\nu\alpha}'$ it follows that:

$$\sum_\alpha \frac{\partial \mathfrak{f}_{\mu\nu\alpha}'}{\partial x'_\alpha} = \frac{1}{\sqrt{g_1}}\sum_{\kappa,\lambda} a_{\kappa\mu}a_{\lambda\nu}\sum_\alpha \frac{\partial \mathfrak{f}_{\kappa\lambda\alpha}}{\partial x_\alpha}.$$

Substituting according to (2), (6), (14):

$$\sum_\alpha \frac{\partial \mathfrak{f}_{\kappa\lambda\alpha}}{\partial x_\alpha} = -\kappa h_{\kappa\lambda} = -2\kappa\frac{\partial H}{\partial g_{\kappa\lambda}},$$

by virtue of (46) and (61) results in:

$$\frac{\partial \mathfrak{f}_{\mu\nu x}'}{\partial x'} + \frac{\partial \mathfrak{f}_{\mu\nu y}'}{\partial y'} + \frac{\partial \mathfrak{f}_{\mu\nu z}'}{\partial z'} + \frac{\partial w_{\mu\nu}'}{\partial t'} = -\kappa h_{\mu\nu}', \qquad h_{\mu\nu}' = 2\frac{\partial H'}{\partial g_{\mu\nu}'}. \tag{64}$$

The definition of the $\mathfrak{W}_\nu'$ and $\mathfrak{B}_\nu'$, (55) and (63) implies:

$$\sum_{\nu}\frac{\partial}{\partial x_{\nu}'}(\mathfrak{W}_{\mu}'\cdot\mathfrak{V}_{\nu}') = \frac{1}{\sqrt{g_1}}\sum_{\lambda}\alpha_{\lambda\mu}\sum_{\nu}\frac{\partial}{\partial x_{\nu}}(\mathfrak{W}_{\lambda}\cdot\mathfrak{V}_{\nu}).$$

From this equation a small calculation according to (38) gives:

$$\left.\begin{aligned}\frac{\partial}{\partial x'}(\mathfrak{W}_{\mu}'\cdot\mathfrak{V}_1')+\frac{\partial}{\partial y'}(\mathfrak{W}_{\mu}'\cdot\mathfrak{V}_2')+\frac{\partial}{\partial z'}(\mathfrak{W}_{\mu}'\cdot\mathfrak{V}_3')-i\frac{\partial}{\partial t'}(\mathfrak{W}_{\mu}'\cdot\mathfrak{V}_4')\\ =\sum_{m,n}\frac{\partial H_e'}{\partial g_{mn}'}\frac{\partial g_{mn}'}{\partial x_{\mu}'}.\end{aligned}\right\} \tag{65}$$

We have now obtained each and every equation of Einstein's theory of gravitation in terms of the transformed quantities; and it turned out that in terms of the latter they have exactly the same form as the original equations in the non-transformed quantities. Considering that the transformed gravitational potential $(g_{\mu\nu}')$ has the scalar value -1 at infinity, whereas the non-transformed $(g_{\mu\nu})$ becomes $(g^1_{\mu\nu})$ at infinity, we see that the transformation property just proved signifies the same as the following theorem:

The Theorem of the Relativity of the Gravitational Potential. *If two empty spaces differ only in that the gravitational potential in one of them has the scalar value* -1, *but in the other an arbitrary tensor value* $g^1_{\mu\nu}$, *then all physical processes in the two spaces proceed in exactly the same fashion, provided that space and time in the first space is specified by means of an ordinary orthogonal system of coordinates* (x, y, z, it), *whereas in the second they are specified by means of a certain oblique system of coordinates* (x', y', z', it') *defined by the eqs. (42) and (48).*

From this it is clearly seen that the "generalized theorem of relativity" plays exactly the same role in Einstein's theory as what I call the "theorem of the relativity of the gravitational potential" (loc. cit. III, p. 61) does in my theory. However, the transformations in Einstein's theory are much more complicated than the extremely simple transformation that is valid in my theory, which is represented by the formulas on p. 63 of my treatise III. This is obvious because calculation with tensors is generally more complicated than with scalars. However, in one respect the difference of Einstein's relativity theorem as opposed to mine is really significant. Whereas only the quantities that specify the state of the aether are transformed in my theory, while the coordinates remain unchanged, in Einstein's theory the coordinates are transformed as well. Therefore the transformations of the two theorems of relativity, that of motion and that of gravitational potential, are very similar to each other in Einstein's theory, and that is probably the reason why Einstein could initially regard his theorem as a generalization of the principle of relativity of motion.[9] From the point of
[172] | view of a physicist the main distinction between the two theories is that by my relativity theorem the laws of nature are not affected at all by the gravitational potential, whereas by Einstein's relativity theorem they are affected as if the units of length and time are changed by the presence of a gravitational potential that differs from -1. So,

according to Einstein's theory the speed of light, the frequencies of spectral lines, the dimensions of atoms and of bodies composed of them, are supposed to change with the gravitational potential, whereas according to my theory nothing of all this should be observable.

8. EQUALITY OF INERTIAL AND GRAVITATIONAL MASS OF CLOSED SYSTEMS IN THE THEORIES OF EINSTEIN AND MIE

Einstein has shown in the report on his lecture held at the Vienna Naturforscherversammlung how one can prove that the two masses of closed systems are equal.[10] If one substitutes the values (32) for h^e_{mn} and (34) for h^g_{mn} into the fundamental eqs. (2) and simultaneously notes (29), these equations become:

$$\sum_\alpha \frac{\partial \mathfrak{k}_{mn\alpha}}{\partial x_\alpha} = -\kappa \frac{\rho}{\sqrt{-g^0_{44}}} \mathfrak{B}_m \mathfrak{B}_n + \kappa \sum_{\mu,\nu,\alpha} \left(\gamma_{\alpha m} \mathfrak{g}_{\mu\nu\alpha} \mathfrak{k}_{\mu\nu n} - \frac{1}{2} \gamma_{mn} \mathfrak{g}_{\mu\nu\alpha} \mathfrak{k}_{\mu\nu\alpha} \right) - \sum_{\mu,\nu,\alpha} \gamma_{m\nu} \frac{\partial g_{\mu\nu}}{\partial x_\alpha} \mathfrak{k}_{\mu n\alpha}.$$

Now take the four equations that correspond to a fixed value of n, as m ranges over the sequence of numbers $1, 2, 3, 4$, multiply each equations by g_{mn}, and add. This yields, with the use of (20):

$$\sum_\alpha \frac{\partial}{\partial x_\alpha} \sum_m g_{mn} \mathfrak{k}_{mn\alpha} = -\kappa \left(\mathfrak{W}_n \mathfrak{B}_n + H_g - \sum_{\mu,\nu} \mathfrak{k}_{\mu\nu n} \mathfrak{g}_{\mu\nu n} \right)$$

or from (39):

$$\sum_\alpha \frac{\partial}{\partial x_\alpha} \sum_m g_{mn} \mathfrak{k}_{mn\alpha} = -\kappa T_{nn}.$$

Similarly by multiplying the four equations one after the other by g_{mp} and adding one obtains:

$$\sum_\alpha \frac{\partial}{\partial x_\alpha} \sum_m g_{mp} \mathfrak{k}_{mn\alpha} = -\kappa \left(\mathfrak{W}_p \mathfrak{B}_n - \sum_{\mu,\nu} \mathfrak{g}_{\mu\nu p} \mathfrak{k}_{\mu\nu n} \right) = -\kappa T_{pn}.$$

---

9 In the introduction to the treatise *Outline of a Generalized Theory of Relativity* etc., Mr. Einstein states the hypothesis "that a homogeneous gravitational field can physically be completely replaced by a state of acceleration of the reference system."[6] Apparently he has the mistaken notion, that this hypothesis (the equivalence hypothesis) is the foundation of the theory developed by him. That would indeed be a more general relativity of motion. In his Vienna lecture Mr. Einstein only demands of the theory of gravitation that the "observable laws of nature do not depend on the absolute magnitude of the gravitational potential" (postulate 4, this journal 14, 1250, 1913) [in this volume]. That is the relativity principle of the gravitational potential, which is what Einstein's theory really satisfies.

10 This journal, 14, 1258, eq. (7b), 1913 [in this volume].

So the following holds in general:

$$\sum_{\alpha}\frac{\partial}{\partial x_{\alpha}}\sum_{m}g_{mp}\mathfrak{k}_{mn\alpha}=-\kappa T_{pn}\,,\quad p=n\quad\text{or}\quad p\neq n. \tag{66}$$

By integrating the eqs. (66) over the volume occupied by the closed system and over a time during which the components of the gravitational as well as the inertial mass of the system do not change, one obtains:

$$g^{1}_{1\mu}m^{g}_{1\nu}+g^{1}_{2\mu}m^{g}_{2\nu}+g^{1}_{3\mu}m^{g}_{3\nu}+g^{1}_{4\mu}m^{g}_{4\nu}=m^{i}_{\mu\nu}. \tag{67}$$

Here $m^{i}_{\mu\nu}$, the (μ,ν) component of the inertial mass, denotes the following integral:

$$m^{i}_{\mu\nu}=\int T_{\mu\nu}dV, \tag{68}$$

and further $g^{1}_{\mu\nu}$ denote the components of the gravitational potential on the boundary of the volume occupied be the closed system. As in Section 7 (p. 169), we assume that the potential $(g^{1}_{\mu\nu})$ can be regarded as constant on the entire boundary surface.

The following equations are easily derived from (67):

$$m^{g}_{\mu\nu}=\gamma^{1}_{\mu1}m^{i}_{1\nu}+\gamma^{1}_{\mu2}m^{i}_{2\nu}+\gamma^{1}_{\mu3}m^{i}_{3\nu}+\gamma^{1}_{\mu4}m^{i}_{4\nu}. \tag{69}$$

Let us for brevity call the component m^{i}_{44} the inertial mass m_i of the system:

$$m_i=\int T_{44}dV. \tag{70}$$

If the system moves through space with velocity $\mathfrak{q}$ it is easy to derive from *Laue's theorem* that

$$\left.\begin{aligned} &m^{i}_{\mu\nu}=-m_i\mathfrak{q}_{\mu}\mathfrak{q}_{\nu},\\ &\mathfrak{q}_1=\mathfrak{q}_x,\quad \mathfrak{q}_2=\mathfrak{q}_y,\quad \mathfrak{q}_3=\mathfrak{q}_z,\quad \mathfrak{q}_4=i.\end{aligned}\right\} \tag{71}$$

And it follows from (69) that:

$$\left.\begin{aligned} &m^{g}_{\mu\nu}=-m_i\mathfrak{p}_{\mu}\mathfrak{q}_{\nu},\\ &\mathfrak{p}_{\mu}=\gamma^{1}_{\mu1}\mathfrak{q}_1+\gamma^{1}_{\mu2}\mathfrak{q}_2+\gamma^{1}_{\mu3}\mathfrak{q}_3+\gamma^{1}_{\mu4}\mathfrak{q}_4.\end{aligned}\right\} \tag{72}$$

The weights of two material bodies that are moving in the same gravitational field with the same velocities are mathematically exactly proportional to their inertial masses.

So we have the theorem to which Einstein still attaches such great importance, once the principle of the identity of the two masses had to be dropped. But from the procedure of the proof it is easy to recognize that this theorem has nothing to do with the actual main assumptions of Einstein's theory, which I mentioned on p. 714

[p. 169 in the original]; that rather it is based mainly on the inessential assumptions that Einstein introduced as supplements into the theory. Most of all, | to prove the theorem one absolutely must adopt the assumption (41): [173]

$$H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}$$

and the assumption (19) or (30):

$$H_g = \frac{1}{2}\sum \mathfrak{g}_{\mu\nu\alpha}\mathfrak{f}_{\mu\nu\alpha}$$

as correct.

The role played in the proof by these two assumptions is recognized most clearly if they are also introduced into the theory of gravitation that I have suggested. In formulating this theory I followed the principle of making no arbitrary auxiliary assumptions if possible, but developing the consequences purely from a single main assumption. This is the assumption that the gravitational mass is completely identical with the rest mass. Certain quite definite reasons speak against introducing the auxiliary assumptions under discussion, as we shall see. Let us, however, temporarily ignore these reasons and adopt both assumptions as correct. So we put:

$$H = H_e + H_g, \quad T_{\alpha\beta} = T^e_{\alpha\beta} + T^g_{\alpha\beta}$$
$$H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}.$$
$$H_g = \frac{1}{2}\mathfrak{g}\mathfrak{f}.$$

In the theory of the scalar potential we then have (loc. cit. III, p. 34, eq. (105)):

$$T^g_{\alpha\alpha} = H_g - \mathfrak{g}_\alpha\mathfrak{f}_\alpha, \quad T^g_{\alpha\beta} = -\mathfrak{g}_\alpha\mathfrak{f}_\beta.$$

A quite simple calculation yields, using the eq. (85) of my treatise III, p. 28:

$$\frac{\partial T^g_{\alpha 1}}{\partial x} + \frac{\partial T^g_{\alpha 2}}{\partial y} + \frac{\partial T^g_{\alpha 3}}{\partial z} - i\frac{\partial T^g_{\alpha 4}}{\partial t} = \frac{\partial H_g}{\partial \omega}\mathfrak{g}_a - \mathfrak{g}_a\left(\frac{\partial \mathfrak{f}_x}{\partial x} + \frac{\partial \mathfrak{f}_y}{\partial y} + \frac{\partial \mathfrak{f}_z}{\partial z} + \frac{\partial w}{\partial t}\right).$$

But we have:

$$\frac{\partial \mathfrak{f}_x}{\partial x} + \frac{\partial \mathfrak{f}_y}{\partial y} + \frac{\partial \mathfrak{f}_z}{\partial z} + \frac{\partial w}{\partial t} = -\kappa h,$$

where h means the density of gravitational mass (eq. (86) loc. cit. III, p. 28) and further (eq. (93) loc. cit. III, p. 30):

$$-\kappa h = \frac{\partial H}{\partial \omega}.$$

If we also split h into two terms $h = h_e + h_g$, where:

$$h_e = -\frac{1}{\kappa}\frac{\partial H_e}{\partial \omega}, \quad h_g = -\frac{1}{\kappa}\frac{\partial H_g}{\partial \omega},$$

then it is easily seen that

$$\frac{\partial T^g_{\alpha 1}}{\partial x} + \frac{\partial T^g_{\alpha 2}}{\partial y} + \frac{\partial T^g_{\alpha 3}}{\partial z} - i\frac{\partial T^g_{\alpha 4}}{\partial t} = \kappa h_e \mathfrak{g}_\alpha.$$

But since:

$$\frac{\partial T_{\alpha 1}}{\partial x} + \frac{\partial T_{\alpha 2}}{\partial y} + \frac{\partial T_{\alpha 3}}{\partial z} - i\frac{\partial T_{\alpha 4}}{\partial t} = 0,$$

it follows from this:

$$\frac{\partial T^e_{\alpha 1}}{\partial x} + \frac{\partial T^e_{\alpha 2}}{\partial y} + \frac{\partial T^e_{\alpha 3}}{\partial z} - i\frac{\partial T^e_{\alpha 4}}{\partial t} = -\kappa h_e \mathfrak{g}_\alpha, \tag{73}$$

an equation that is the exact analogue of eq. (38) of Einstein's theory. We can regard (73) as the equation of motion of a particle having the inertial mass

$$m_i = \int T^e_{44} dV \tag{74}$$

under the influence of the gravitational force in addition to the forces that correspond to the state variables of the aether occurring in the $T^e_{\alpha\beta}$. The gravitational mass of a particle is to be reckoned as:

$$m_g = \int h_e dV. \tag{75}$$

Now I introduce the main assumption on which my theory is based:

$$\left.\begin{aligned} h &= H \\ h_e &= H_e, \quad h_g = H_g. \end{aligned}\right\} \tag{76}$$

Following eq. (41) I put:

$$h_e = H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44},$$

so that, if for simplicity we assume the body to be at rest, *Laue's theorem* implies:

$$\left.\begin{aligned} &\int h_e dV = \int T^e_{44} dV \\ &m_g = m_i. \end{aligned}\right\} \tag{77}$$

That is because Laue's theorem is separately applicable to each of the two terms in the energy tensor $T^e_{\alpha\beta}$ and $T^g_{\alpha\beta}$, since in the interior of a complete stationary system each of the two components of the energy current $T^e_{4\beta}$ and $T^g_{4\beta}$ must vanish separately.

If Einstein's auxiliary assumptions $H_e = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}$ *and* $H_g = \frac{1}{2}\mathfrak{g}\mathfrak{f}$ *were to be introduced into the theory of gravitation suggested by me, then also in this theory the weights of two bodies that move with the same velocity in the same gravitational field would be proportional to their inertial masses.*

The theorem about the equality of the two inertial masses of closed systems is not at all a consequence of the two main assumptions of Einstein's theory of gravitation, the assumption of a tensor potential and the assumption of a peculiar transformation property of the basic equations; but it follows | *from the inessential, incidental auxiliary assumptions of the theory.* [174]

9. INTERNAL CONTRADICTION IN EINSTEIN'S AUXILIARY ASSUMPTIONS

Because the trace $T_{11} + T_{22} + T_{33} + T_{44}$ is a four-dimensional scalar, transformation to a coordinate system, in which the closed material system under consideration is at rest, yields:

$$\begin{aligned}&\int (T_{11} + T_{22} + T_{33} + T_{44})\,dV \\ &= \int (T_{11}^{\ 0} + T_{22}^{\ 0} + T_{33}^{\ 0} + T_{44}^{\ 0})\sqrt{1-\mathfrak{q}^2}\,dV_0 \\ &= \sqrt{1-\mathfrak{q}^2}\int T^0_{44}\,dV_0,\end{aligned}$$

where $\mathfrak{q}$ denotes the velocity with which the material system moves in the original coordinate system. Because T^0_{44} equals the denseness of the rest energy, that is H, it follows that:

$$\int (T_{11} + T_{22} + T_{33} + T_{44})\,dV = \int H\,dV. \tag{78}$$

But by assumption (41):

$$T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44} = H_e,$$

it follows from (78) that:

$$\int (T^g_{11} + T^g_{22} + T^g_{33} + T^g_{44})\,dV = \int H_g\,dV. \tag{79}$$

Equation (79) is in direct contradiction with the auxiliary assumption (19). Namely, if one writes this auxiliary assumption in the form (30) then (39) results in:

$$T^g_{11} + T^g_{22} + T^g_{33} + T^g_{44} = 2H_g. \tag{80}$$

The two auxiliary assumptions (19) and (41) of Einstein's theory are mutually contradictory.

It is remarkable that these two assumptions are necessary precisely for the proof of the theorem of the equality of the two masses. It should probably not be difficult to eliminate the internal contradiction from Einstein's theory. *However one may well suppose that the removal of the internal contradiction will be accompanied by the failure of the theorem of the equality of the two masses*.

From the general theoretical investigations that I made concerning the nature of matter one can discern that assumption (41) is by itself untenable, even apart from the contradiction with assumption (19). I can say that this realization was indeed the reason for me to abandon *ab initio* the theorem about the equality of the two masses of a closed system; or else considerations such as those presented in Section 8 would rather quickly have come to mind.

APPENDIX

10. NORDSTRÖM'S TWO THEORIES OF GRAVITATION

Mr. Gunnar Nordström has published two different theories of gravitation, both of which he obtained by suitable modifications of Abraham's equations of gravitation (which are not in accord with the principle of relativity). The first of these was first published by him toward the end of the year 1912.[11] There the rest density of energy is decomposed into three terms:

$$H = H_e + H_p + H_g,$$

of which the second depends only on the elastic tensions of matter, the third only on the field strength of gravitation, and the first on all the remaining state variables. Mr. Nordström calls H_e in particular the rest density of the matter's inertial mass and puts:

$$\frac{\partial H}{\partial \omega} = -\kappa H_e.$$

Accordingly in this theory of Nordström's we have:

$$H = e^{-\kappa\omega} H_e' + H_p + H_g,$$

where H_e' as well as H_p and H_g no longer depend on the gravitational potential ω.

Evidently this theory is rather similar to the one suggested by me. I have:

$$H = e^{-\kappa\omega} H',$$

11 This journal 13, 1126, 1912; *Ann. d. Phys*. 40, 856, 1913 [both in this volume].

where H' no longer depends on ω, so I have avoided the somewhat artificial decomposition of H into three terms.

Nordström's notation deviates strongly from mine, one has to put:

$$H_e = \nu, \quad \kappa = \frac{g}{c^2}, \quad \omega = -\Phi, \quad \mathfrak{g}_x = \mathfrak{k}_x = -\frac{\partial \Phi}{\partial x} \quad \text{etc.},$$

in order to obtain Nordström's equations, in addition one must note that I have set the speed of light in an ideal vacuum equal to 1, which Nordström calls c.

Nordström's second theory[12] appeared only recently. Incidentally, this is the theory about which Mr. Einstein spoke in his Vienna lecture,[13] whereas in the | discussion I meant the older theory, the only one published to that date. The second theory [175] contains the two assumptions, that H is to be split into two terms:

$$H = H_e + H_g,$$

of which only the second depends on the field strength of gravity $(\mathfrak{g}, iu)$, and that:

$$H_g = \frac{1}{2}(\mathfrak{g}^2 - u^2).$$

The gravitational potential ω shall reside only in H_e, as in Nordström's first theory. A further assumption is made about the density of inertial mass h (loc. cit. eqs. (1), (2), (14), (15)):

$$h = \frac{\nu}{1 - \kappa\omega}$$

$$\nu = T^e_{11} + T^e_{22} + T^e_{33} + T^e_{44}.$$

Finally (loc. cit. eq. (27)) the quantity:

$$\frac{\nu}{(1 - \kappa\omega)^2}$$

shall not depend on ω.

However, these assumptions incorporate a grave internal contradiction. Namely, by noting, as I proved in my treatise III on p. 30, that the energy principle can be satisfied only if

$$-\kappa h = \frac{\partial H}{\partial \omega}$$

---

12 *Ann. d. Phys.* 42, 533, 1913 [in this volume].

13 In his lecture *On the Present State of the Problem of Gravitation*, Mr. Einstein mentioned of all theories other than his own only this second theory of Nordström's. The comments on Abraham's theory found in the report in this journal 14, 1250, 1913 [in this volume] did not come up in the lecture itself. I want to mention this here in order to explain my remarks at the beginning of the discussion (this journal 14, 1262, 1913).

or, since Nordström's H_g is independent of ω:

$$-\kappa h = \frac{\partial H_e}{\partial \omega}$$

one sees that Nordström's assumptions imply:

$$\frac{1}{(1-\kappa\omega)^3}\frac{\partial H_e}{\partial \omega} = -\kappa H_e' = -\kappa\frac{T^e_{11}+T^e_{22}+T^e_{33}+T^e_{44}}{(1-\kappa\omega)^4}$$

where H_e' is to denote an ω-independent quantity. Integration results in:

$$H_e = \frac{1}{4}(1-\kappa\omega)^4 H_e' = \frac{1}{4}(T^e_{11}+T^e_{22}+T^e_{33}+T^e_{44}).$$

We therefore have:

$$\int H_e dV = \frac{1}{4}\int(T^e_{11}+T^e_{22}+T^e_{33}+T^e_{44})dV,$$

where the integral is to be taken over the volume of a closed system.

But according to Laue's theorem we must have:

$$\int H_e dV = \int(T^e_{11}+T^e_{22}+T^e_{33}+T^e_{44})dV.$$

Accordingly the assumptions made by Mr. Nordström must somehow lead to contradictions with the energy principle, which of course must not happen. I have not further explored whether and how this error can be eliminated from Nordström's ansatz, and therefore I do not want to discuss this theory here any further, although I believe that it would be quite interesting when consistently developed.

The notation is the same as in Nordström's older papers. In some formulas Φ is replaced by $\Phi - \Phi_0$, for example:

$$1-\kappa\omega = 1+\frac{g}{c^2}(\Phi-\Phi_0)$$

and further he sets:

$$\frac{\kappa}{1-\kappa\omega} = \frac{g}{1+\frac{g}{c^2}(\Phi-\Phi_0)} = g(\Phi)$$

as well as:

$$\frac{1-\kappa\omega}{\kappa} = \frac{c^2}{g}+\Phi-\Phi_0 = \Phi'.$$

11. CONCLUSION: SUMMARY

In his Vienna lecture Mr. Einstein was very articulate about the ultimate goal of his investigations.

1. In his research an attempt is made to enlarge the theory of relativity; in particular the principle of relativity, which at first is valid only for uniform motion, is to be extended to accelerated motion, at least to uniform acceleration. As Mr. Einstein himself emphasized, this amounts to demanding covariance of the laws of nature not only with respect to linear substitutions, but also with respect to nonlinear substitutions.

2. The generalization of the principle of relativity is to be achieved by allowing the complete replacement of the accelerated motion of a material system by a gravitational field. As Mr. Einstein puts it, a physicist from his standpoint can characterize the gravitational field as "fictitious," because a suitable transformation of the basic equations of physics can always make the gravitational field disappear at the location in question, by replacing it with an equivalent state of acceleration. Conversely one can of course equally well designate an acceleration of the system as fictitious. This hypothesis of the equivalence of gravitational field and acceleration is of course realizable only if inertial and gravitational mass are identical in their nature.

Even acknowledging the extremely ingenious and painstaking workmanship that Mr. Einstein has devoted to the achievement of the stated goal, one can nevertheless say nothing more than that his attempt has had only a negative result. |

1. The degree of generalization of the relativity principle achieved in Einstein's [176]
work concerns only linear transformations, so it has nothing whatever to do with accelerated motion. In the present analysis I have demonstrated that this "generalization" means nothing more than that besides the relativity of motion there exists a relativity of the gravitational potential. This second relativity is valid in my theory as well, and there it is in fact achieved by extremely simple means.

2. The equivalence hypothesis seems to me untenable already for this reason, that there is no such thing as the identity of inertial and gravitational mass in Einstein's theory. Certainly, introducing several auxiliary assumptions produces the proposition that the gravitational and inertial mass of closed systems are strictly proportional to each other. But this proposition is by no means a consequence of the transformation properties of the fundamental equations, as it ought to be according to the equivalence hypothesis, it would be equally valid in my theory if the auxiliary assumptions just mentioned were to be imported also into it. Further, it has turned out that these auxiliary assumptions contain an internal contradiction, and thereby the proposition of the equality of the two masses becomes untenable even in the modest form it eventually assumed.

As a positive result of the present investigation I count the demonstration that in any theory in which gravitational mass is a four-dimensional tensor, an identity of the tensor of inertial mass with the tensor of gravitational mass is impossible, come what may. As far as I can see this surely establishes quite generally that a principle of the identity of the two masses cannot be valid. Whether one can attain from this a general

demonstration of the impossibility of Einstein's equivalence hypothesis cannot be said without more detailed investigation, but to me it seems quite possible. In any event I am inclined to believe that the failure of Einstein's attempt is to be explained by the impossibility of success. In the discussion at Einstein's lecture I have pointed out that a generalization of the relativity principle as intended by Einstein will probably always lead to contradictions with the general principles of inquiry in physics (this journal 14, 1264). Now it would be interesting if the impossibility of generalization could be demonstrated from a different point of view by rigorous mathematics. In this context a proposition announced by Mr. Einstein (this journal 14, 1257) seems to me significant: according to it no system of fundamental equations can be devised that would be covariant in their entirety for arbitrary substitutions.

EDITORIAL NOTES

[1] *Dichtigkeit* is translated as "denseness," in order to respect the distinction Mie draws between the tensor quantity *Dichtigkeit* and the scalar *Dichte* ("density").

[2] Mie uses the term *Wesensgleichheit*, translated as "unity of essence," alluding to Einstein's use of this term to describe the relation between inertia and gravitation.

[3] In the original, Mie mistakenly refers to eq. (16) rather than eq. (17).

[4] In eq. (23), the subscript $_\lambda$ of the first occurrence of $g_{1\lambda}$ is missing in the original.

[5] In the original text the summation in the second line in the following equations runs over λ; here it has been corrected to κ.

[6] Here, Mie leaves out Einstein's qualification that the gravitational field is infinitesimally extended.

GUSTAV MIE

THE PRINCIPLE OF THE RELATIVITY OF THE GRAVITATIONAL POTENTIAL

Originally published as "Das Prinzip von der Relativität des Gravitationspotentials" in Arbeiten aus den Gebieten der Physik, Mathematik, Chemie. Festschrift Julius Elster und Hans Geitel zum sechzigsten Geburtstag, Braunschweig: Friedr. Vieweg & Sohn 1915, pp. 251–268. Author's note: Greifswald (Physical Institute of the University) April 27, 1915.

1. Gravity has long eluded theoretical investigation, and, apart from the meager knowledge gained by experience, the main reason is that the gravitational field exhibits the peculiarity that the field itself is amplified when it performs external work. It is therefore difficult to design the theory so that it does not conflict with the energy principle. Admittedly the magnetic field of two current-carrying conductors shows the same peculiarity. However, in that case one readily recognizes in the apparatus that provides the current the source of energy both for the energy increase of the magnetic field and for the energy carried off as work. For the gravitational field such an external energy source is absent, and therefore one formerly used to assume in gravitational theories that the energy of the field is negative, so that upon amplification of the field a positive energy is released and is manifested as work gained. Every theory of gravitation that is built upon the scheme of Maxwell's equations must make the assumption of a negative energy. But this assumption is untenable, because a field whose energy is negative cannot be in stable equilibrium, but is always unstable. Namely, whereas an electrostatic field, for example, exhibits that distribution of lines of force for which the energy has the smallest possible value, a field that is similarly constituted but with negative energy has of course precisely the largest possible value at equilibrium. It is therefore to be expected that when the equilibrium is slightly perturbed, the field will continually release energy to the exterior | while simultaneously [252]
moving further and further away from its equilibrium state. So in this way one does not attain a satisfactory theory of gravity.

A simple, feasible way leading out of this difficulty was first indicated by M. Abraham.[1] This way consists of including in the state variables, upon which the

1 M. Abraham, *Ann. d. Phys.* 38, 1056 (1912).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

amount of energy per cubic centimeter depends, the *potential* of the field, formerly taken to be only a "mathematical construct", rather than only the *field strength* of the gravitational field, formerly considered exclusively. Now, when two gravitating masses approach each other, there is on the one hand an increase in the energy of the field, counted as positive exactly like that of an electric field of similar appearance, but at the same time there is a change, a decrease to be exact, in the internal energy of the approaching material bodies, because in them the potential of the gravitational field becomes different. So the two gravitating masses release a part of their internal energy as a consequence of the change in their gravitational potentials, and thus provide the source for the work gained due to the attraction as well rather than only the increase in energy of the gravitational field. Thus, in the case of the gravitational field, matter under the influence of a changed potential performs the same task as the current source in the case of the magnetic field between current-carrying conductors mentioned above.

However, it is important to note that this procedure cannot be carried through for a field whose potential is a four vector, like that of the electromagnetic field. Therefore M. Abraham has derived the gravitational field from a potential that is invariant under Lorentz transformations, so it is a four dimensional scalar. This results in a theory without further difficulties. From the investigations of Messrs. A. Einstein and M. Grossmann,[2] it follows that one can achieve the same by deriving the gravitational field from a potential that is a four dimensional tensor. However, the theory of the tensorial gravitational potential is significantly more complicated than the scalar one, and since in spite of its complications it does not exhibit any advantages whatever over the theory of a scalar potential, I prefer to stay with the scalar potential. |

[253] Thus the potential plays a very important role in the theory of gravitation. It may be concluded from my investigations on the theory of matter[3] that the four potential of the electromagnetic field, no less than the scalar potential of the gravitational field, has to be counted among the state variables on which the energy depends. But this dependence must be such that when the electromagnetic potential is changed in the region where a material particle is located, the *net* change in energy of the particle is either nil or an infinitely small amount of higher order. For the work due to the action of electromagnetic forces is obtained with great accuracy as equal to the sum of the energy change experienced by the field due to displacement of the bodies that generate the field and the energy provided by the sources of electricity used in doing this. Therefore the energy of the material particles is to be regarded as constant, independent of the field strength and the potential prevailing in their vicinity. So it has been possible to develop a theory of the electromagnetic phenomena, adequate for a large class of empirical facts, in which the four potential does not occur except as a purely mathematical construct. It is different in the case of gravity. The total energy of the

2 A. Einstein und M. Grossmann, *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig, B. G. Teubner, 1913.

3 G. Mie, *Ann. d. Phys., Abhandl.* I, 37, 511 (1912); II, 39, 1 (1912); III, 40, 1 (1913).

material particles depends on the gravitational potential prevailing in their surroundings to such an extent that the sign of the action of the force is thereby reversed. Accordingly it seems that in order to set up a theory of gravitation one must also specify how the gravitational potential enters into the physics of the aether. For this reason—in contrast to the potential-free theory of electromagnetism—countless theories of gravitation are possible, which all show the same form of the basic equations and differ only in the way that the gravitational potential enters into the expression for the energy. Indeed several theories of that kind have already been proposed, but one could of course add to them any number of others. But this procedure is hardly satisfactory because following it one can never completely avoid pulling quite arbitrary assumptions out of thin air. And this without knowing anything empirical about how the gravitational potential enters into the expression for the energy! Thus the consequences | derived from the theories so obtained will also be given little credibil- [254]
ity. Therefore in the following I want to investigate how far one can get without arbitrary specializations, presupposing as correct only a few quite general principles, which have a certain inherent probability due to their simplicity.

2. The first principle upon which I base the theory is *the principle of relativity*. By this I mean the proposition that all basic equations of the physics of the aether admit the Lorentz transformation. This principle has been accepted in all newer gravitational theories except that of Abraham. Why Mr. Abraham considers his special assumptions more important than the principle of relativity is something I cannot fathom.

3. Secondly I presuppose *Hamilton's principle*. The basic equations of the physics of the aether, whatever their detailed appearance, can at any rate always be divided into two groups, so that exactly as many state variables occur in each of the two groups of equations as are necessary and sufficient for a complete description of the state of the aether. So, for this description one can choose at will either the variables of the first group of equations or equally well those of the second group. In my papers I have differentiated between the two types of state variables as intensive and extensive quantities [*Intensitätsgrößen* and *Quantitätsgrößen*]. The variables of the two groups can be coordinated with each other into pairs of conjugate [*entsprechende*] variables. Hamilton's principle amounts in essence to the proposition that the state variables of one group can be calculated from those of the other group if one knows only a single function of them, the Hamiltonian. To calculate a desired state variable from given variables of the other group one has to take the partial derivative of the Hamiltonian with respect to the conjugate state variable; this partial derivative is the desired quantity. In most cases it turns out to be advantageous to consider the intensive quantities as the primary state variables. I have always denoted the Hamiltonian function of the intensive quantities by Φ and have called it "the world function."

If Hamilton's principle is valid, the equations of motion of mechanics and the energy principle can easily be derived from it as consequences. If one prefers not to presuppose it, then it | is at least highly questionable whether the energy principle can [255]
be maintained. In the gravitational theory of G. Nordström[4] the special assumptions

are chosen in such a way that Hamilton's principle is not valid, so that no world function exists. It is incomprehensible why Mr. Nordström prefers his special assumptions over Hamilton's extraordinarily clear and simple principle.[5]

Concerning the variables that are supposed to determine the world aether, and on which the world function Φ will accordingly depend, we shall assume the guiding principle that we shall try to get by with as few variables as possible. We can safely leave the question of whether it will prove necessary as science progresses to increase the number of state variables undecided; probably the propositions that we can derive from our general principles will not be significantly modified by this. At the present state of science the following mutually independent variables of state (intensive quantities) suffice: 1. electric field strength $\mathfrak{e}$; 2. magnetic induction $\mathfrak{b}$; 3. electromagnetic four potential $\varphi, \mathfrak{f}$; 4. four vector of the field strength of gravity $\mathfrak{g}, \gamma$; 5. gravitational potential ω. The world function is therefore: $\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}, \mathfrak{g}, \gamma, \omega)$.

The corresponding extensive quantities are: 1. electric displacement $\mathfrak{d}$; 2. magnetic field strength $\mathfrak{h}$; 3. electric charge and electric current $\rho, \mathfrak{v}$; 4. four vector of excitation of the gravitational field $\mathfrak{k}, \chi$; 5. density of gravitational mass h.

Hamilton's principle leads to the following equations:

$$\mathfrak{d} = -\frac{\partial \Phi}{\partial \mathfrak{e}}, \mathfrak{h} = \frac{\partial \Phi}{\partial \mathfrak{b}}, \rho = \frac{\partial \Phi}{\partial \varphi}, \mathfrak{v} = -\frac{\partial \Phi}{\partial \mathfrak{f}}, \mathfrak{k} = \frac{\partial \Phi}{\partial \mathfrak{g}}, \chi = -\frac{\partial \Phi}{\partial \gamma}, h = -\frac{\partial \Phi}{\partial \omega}. \qquad (1)$$

[256] | Here I generally use the same notation as in my earlier papers (cf. Theory of Matter III, p. 30). Except that previously I chose the less practical notation u, w in place of the letters γ, χ; further instead of h I previously used γH or (Physik. Zeitschr. 15, 175 (1914)) κh, where γ resp. κ means the gravitational constant $1,016 \cdot 10^{-24}$. Let us denote the gravitational mass of the particle by m_g, that is:

$$m_g = \int h dV, \qquad (2)$$

where the integral is to be taken over the entire volume occupied by the particle (cf. Theory of Matter III p. 6); then the force $\mathfrak{P}$ acting on it in a field $\mathfrak{g}$ is:

$$\mathfrak{P} = m_g \mathfrak{g} \ . \qquad (3)$$

4. My third assumption is *the principle of the relativity of the gravitational potential*.

---

4 G. Nordström, *Ann. d. Phys.* 42, 533 (1913).

5 In answer to the objection I raised (*Physik. Zeitschr.* 15, 175 (1914)), that then the energy principle must also fail, Mr. G. Nordström has recently tried to show (*Physik. Zeitschr.* 15, 375 (1914)) that in spite of the disagreement with Hamilton's principle his theory does not have to conflict with the energy principle. However, the proof of this has not yet really been established, because Mr. Nordström has not actually put down the basic equations of the theory, but he has only indicated how one might set them up. It seems to me by no means certain that one can proceed according to his indications without using Hamilton's principle, for Mr. Herglotz, who sets up basic equations in his Mechanics of Continuous Media (*Ann. d. Phys.* 36, 493 (1911)) of the type indicated by Mr. Nordström, has certainly considered it necessary to base his investigations on Hamilton's principle.

Introducing the potentials as independent state quantities leads to a peculiar difficulty. It forces us to assume that the properties of matter and the laws of material processes depend on the potentials that prevail at the location where these properties and processes are observed. On the other hand no one has ever been aware of such an influence of the potentials, and if it exists at all it must at least be quite insignificant. Otherwise, although it may have seldom been looked for,[6] one should think that it would already have been noticed on other occasions. So we are confronted with the dilemma that on the one hand the theory absolutely demands an influence of the potentials on physical processes, and that on the other hand experience negates this influence to such an extent that it has become second nature to regard the potentials as purely mathematical, calculational constructs.

This dilemma can be eliminated, initially for the gravitational potential, using the principle at hand in a way that is as simple as it is perfect. The principle declares: |

In two regions of different gravitational potential exactly the same processes can [257]
run according to exactly the same laws if one only thinks of the units of measurement as changing in a suitable way with the value of the gravitational potential.

I shall show that this principle can be realized even as I obtain its mathematical formulation. We assume that the gravitational potential ω has a value Ω that *differs from zero* in an ideal vacuum, at an infinite distance from any matter. Ω is a universal constant of the aether, like the speed of light, the constant of gravity etc.; a Lorentz transformation does not change its value because ω is a four dimensional scalar. I note incidentally that by contrast the four potential of the electromagnetic field $(\varphi, \mathfrak{f})$ must be zero in a vacuum. For, if it had a non-zero value there, this would change upon any Lorentz transformation. One would then have universal constants that would depend on the choice of the spacetime coordinate system, in other words the principle of relativity would not be strictly valid. In the same way the field strengths $(\mathfrak{b}, -i\mathfrak{e})$ and $(\mathfrak{g}, i\gamma)$ must of course also vanish in a pure vacuum. Let us now transform all quantities of state in such a way that each is multiplied by a constant factor:

$$\left.\begin{aligned} \omega &= a \cdot \omega', & (\mathfrak{g}, i\gamma) &= b \cdot (\mathfrak{g}', i\gamma'), \\ (\mathfrak{f}, i\varphi) &= c \cdot (\mathfrak{f}', i\varphi'), & (\mathfrak{b}, -i \cdot \mathfrak{e}) &= d \cdot (\mathfrak{b}', -i\mathfrak{e}') \end{aligned}\right\} \tag{4}$$

The primed and unprimed quantities then differ only by the measurement units. If the same equations are to hold in the primed quantities as in the unprimed ones, then it is absolutely necessary that the world function Φ also experiences no changes through the introduction of new measurement units other than a constant factoring out:

$$\Phi(\mathfrak{e}, \mathfrak{b}, \varphi, \mathfrak{f}, \mathfrak{g}, \gamma, \omega) = e \cdot \Phi(\mathfrak{e}', \mathfrak{b}', \varphi', \mathfrak{f}', \mathfrak{g}', \gamma', \omega') = e \cdot \Phi' . \tag{5}$$

---

6 Some time ago with Prof J. Herweg I have used a good echelon grating to observe the spectrum of a mercury arc lamp, in which quite large values of the magnetic vector potential could be produced by means of nearby electric ring magnets, without generating a magnetic field in the lamp. No trace of an influence of the vector potential on the spectral lines was revealed.

For then, but only then, is it true for the state variables of the other set that they are transformed in the same way:

$$\left.\begin{aligned} h &= \frac{e}{a}\cdot h', & (\mathfrak{k}, i\chi) &= \frac{e}{b}\cdot(\mathfrak{k}', i\chi'),\\ (\mathfrak{v}, i\rho) &= \frac{e}{c}\cdot(\mathfrak{v}', i\rho'), & (\mathfrak{h}, -i\cdot\mathfrak{d}) &= \frac{e}{d}(\mathfrak{h}', -i\cdot\mathfrak{d}') \end{aligned}\right\} \tag{6}$$

[258] | If one now also puts:

$$x = \frac{a}{b}\cdot x', \quad y = \frac{a}{b}\cdot y', \quad z = \frac{a}{b}\cdot z', \quad t = \frac{a}{b}\cdot t', \tag{7}$$

and if the condition:

$$\frac{a}{b} = \frac{c}{d} \tag{8}$$

is satisfied, then, as one can prove easily, the basic equations of the physics of the aether are equally as valid in the primed quantities as in the unprimed ones. If we choose the value $\omega'_\infty = \Omega$, then we have thereby completely reduced all physical problems in a region of gravitational potential $\omega_\infty = \Omega_1 = a\Omega$ to the corresponding problems in an ideal vacuum where the gravitational potential is Ω; that is, the processes in the two regions differ only in the difference of the units of measurement.

This shows that the principle of relativity formulated above is valid if and only if the world function Φ has the property demanded by equation (5). But equation (5) is the condition that Φ is a homogeneous function of the variables.

We can state quite generally that the principle of the relativity of the gravitational potential is identical with the demand that the world function Φ is a homogeneous function of the variables:

$$\omega, \quad (\mathfrak{g}^\kappa, \gamma^\kappa), \quad (\mathfrak{f}^\lambda, \varphi^\lambda), \quad (\mathfrak{b}^\mu, \mathfrak{e}^\mu),$$

where κ, λ, μ denote arbitrary positive or negative, integral or fractional numbers.

I will call the degree of this homogeneous function ν. In the equations (4) and (5) we then have to put:

$$b = a^{\frac{1}{\kappa}}, \quad c = a^{\frac{1}{\lambda}}, \quad d = a^{\frac{1}{\mu}}, \quad c = a^\nu. \tag{9}$$

If we further put (taking note of 8):

$$1 - \frac{1}{\kappa} = \frac{1}{\lambda} - \frac{1}{\mu} = \alpha, \tag{10}$$

then we have:

$$x = a^\alpha\cdot x', \quad y = a^\alpha\cdot y', \quad z = a^\alpha\cdot z', \quad t = a^\alpha\cdot t'. \tag{11}$$

As an example of a theory in which this relativity principle is valid I mention | the [259]
theory of gravitation developed by me on a previous occasion (Theory of Matter III, p. 25ff). To obtain this, one has to specialize by setting:

$$\kappa = 1, \quad \lambda = \mu = \frac{1}{\nu},$$

and further:

$$\Omega = -\frac{\nu}{\gamma},$$

where $\gamma = 1,016 \cdot 10^{-24}$ signifies the gravitational constant. Because ω becomes zero at infinity in the notation chosen previously by me, one has to substitute $(\omega - \Omega)$ everywhere in place of ω in the formulae of my treatise III, in order to adjust them to my current notation. On then obtains the equations of treatise III, if one makes ν infinite. Namely:

$$\lim_{\nu = \infty} \left(\frac{\omega^\nu}{\Omega^\nu} \right) = e^{-\gamma \cdot (\omega - \Omega)}.$$

If we want to transform to a space having $\Omega_1 = \Omega + \omega_0$, then we have to put:

$$a = \frac{\Omega_1}{\Omega},$$

and we have:

$$\lim_{\nu = \infty} a = 1, \quad \lim_{\nu = \infty} (a^\nu) = e^{-\gamma \cdot \omega_0},$$

from which the transformation equations presented in treatise III on p. 63 follow directly.

Finally, equation (10) implies: $\alpha = 0$, $a^\alpha = 1$, so the units of length and time measurements are not changed by the transformation.

Accordingly the theory I developed previously is indeed a special case, or rather a limiting case, of theories in which the principle of relativity of the gravitational potential holds.

5. The only empirical fact that we know about gravitation to date is the *proportionality of the gravitational and inertial mass of a body*. It is interesting that this fact can also be obtained theoretically, as I will now show, by assuming the principle of relativity of the gravitational potential to be correct. |

I focus on a material body that is a complete system in the sense of Laue's theo- [260]
rem.[7] For the sake of generality I assume that the elementary particles of the body execute arbitrary random motion, whereas the body as a whole is at rest. I will mark

7 M. v. Laue, *Das Relativitätsprinzip*, 2nd Ed. p. 208. Braunschweig 1913. In the following I use the formulas developed in *Theory of Matter* III, section 27 and 28 (p. 5–11) and 43 (p. 42, 43) [in this volume].

the time average of state variables by horizontal bars above the respective mathematical symbols, as in my earlier investigations. The principle of relativity of the gravitational potential, according to which Φ is a homogeneous function of the state variables (cf. p. 257) yields

$$\nu \cdot \bar{\Phi} = \overline{\omega \cdot \frac{\partial \Phi}{\partial \omega}} + \frac{1}{\kappa} \cdot \left(\overline{\mathfrak{g} \cdot \frac{\partial \Phi}{\partial \mathfrak{g}}} + \overline{\gamma \cdot \frac{\partial \Phi}{\partial \lambda}} \right) + \frac{1}{\lambda} \cdot \left(\overline{\mathfrak{f} \cdot \frac{\partial \Phi}{\partial \mathfrak{f}}} + \overline{\varphi \cdot \frac{\partial \Phi}{\partial \varphi}} \right) + \frac{1}{\mu} \cdot \left(\overline{\mathfrak{e} \cdot \frac{\partial \Phi}{\partial \mathfrak{e}}} + \overline{\mathfrak{b} \cdot \frac{\partial \Phi}{\partial \mathfrak{b}}} \right),$$

or, using the relations (1):

$$\nu \bar{\Phi} = -\overline{\omega \cdot h} + \frac{1}{\kappa} \cdot (\overline{\mathfrak{g} \cdot \mathfrak{k}} - \overline{\gamma \cdot \chi}) + \frac{1}{\lambda} (\overline{\varphi \cdot \rho} - \overline{\mathfrak{f} \cdot \mathfrak{v}}) - \frac{1}{\mu} \cdot (\overline{\mathfrak{e} \cdot \mathfrak{d}} - \overline{\mathfrak{b} \cdot \mathfrak{h}}). \tag{12}$$

Let the inertial mass of the body be m, the gravitational mass m_g, then we have by equation (2) and according to the Theory of Matter III, equation (116) on p. 42:

$$m_g = \int \bar{h} \cdot dV \tag{13}$$

$$m = \int (\bar{\Phi} + \overline{\gamma \cdot \chi}) \cdot dV. \tag{14}$$

In addition we need the following equations from the *Theory of Matter* III, p. 7: equations (64) and (65), as well as p. 43: equation (117) and (118):

$$\int \overline{\mathfrak{e} \cdot \mathfrak{d}} \cdot dV = \int \overline{\varphi \cdot \rho} \cdot dV \tag{15}$$

$$\int \overline{\mathfrak{b} \cdot \mathfrak{h}} \cdot dV = \int \overline{\mathfrak{f} \cdot \mathfrak{v}} \cdot dV \tag{16}$$

$$\int (\overline{\mathfrak{g} \cdot \mathfrak{k}} - \overline{\gamma \cdot \chi}) \cdot dV = \int \overline{\omega \cdot h} \cdot dV - \omega_\infty \cdot m_g \tag{17}$$

$$3m = \int (\overline{\mathfrak{g} \cdot \mathfrak{k}} + 3 \cdot \overline{\gamma \cdot \chi} - \overline{\mathfrak{e} \cdot \mathfrak{d}} + \overline{\mathfrak{b} \cdot \mathfrak{h}}) \cdot dV. \tag{18}$$

I have everywhere substituted γ and χ for the letters u and w used previously, also h in place of γH, finally I substituted $\omega - \omega_\infty$ for ω. |

[261] If one now notes equation (10):

$$1 - \frac{1}{\kappa} = \frac{1}{\lambda} - \frac{1}{\mu} = \alpha,$$

by a simple calculation one finds from equations (12) to (18):

$$(\nu + 3\alpha) \cdot m - (\nu + 4\alpha) \cdot \int \sqrt{\gamma \cdot \chi} \cdot dV = -\omega_\infty \cdot m_g. \tag{19}$$

In an ideal vacuum, where $\omega_\infty = \Omega$, we have:

$$(\nu + 3\alpha) \cdot m - (\nu + 4\alpha) \cdot \int \sqrt{\gamma \cdot \chi} \cdot dV = -\Omega \cdot m_g. \tag{20}$$

To begin with, focus on the case that the elementary particles of the body are all at rest:

$$\gamma = -\frac{\partial \omega}{\partial t} = 0.$$

Therefore, we have in a region of gravitational potential ω_∞:

$$(\nu + 3\alpha) \cdot m = -\omega_\infty \cdot m_g, \tag{21}$$

and in an ideal vacuum:

$$(\nu + 3\alpha) \cdot m = -\Omega \cdot m_g. \tag{22}$$

Thus the ratio m_g / m is a universal constant.

For a material body whose elementary particles are motionless, the law of the proportionality of inertial and gravitational mass holds with mathematical precision.

If one wished to regard the law of the proportionality of the two masses as a kind of axiom, which must also be satisfied with mathematical precision when the elementary particles of the body execute hidden motions, then one would have to assume, in addition to the principle of the relativity of the gravitational potential, the validity of the relation:

$$\nu + 4\alpha = 0.$$

But this law certainly states only an empirical fact, and even if it is true to very high accuracy according to the experiments of Eötvös, there is no sensible reason why one should accord it a character other than an empirical, approximate one. Even Newton's laws of motion, though treated almost as axioms for hundreds of years, are only approximate propositions according to the theory of relativity. These laws, it is true, are valid to such accuracy that usually one cannot experimentally substantiate any deviations from them. They approach the truth so closely only because the experimentally attainable speeds of material bodies can be characterized as infinitesimal compared | to the speed of light. It would be quite possible that the whole validity of [262]
the law of proportionality of the two masses has a quite similar reason; namely, that the speeds of the hidden motions of the elementary particles of a material body are in general infinitely small compared to the speed of light. In an earlier paper I have tried to estimate what the order of magnitude of the deviation from proportionality of the two masses, due to the thermal motion of the molecules, might be; and I found that even at temperatures of several thousands of degrees Celsius the deviation lies below an experimentally detectable magnitude (*Theory of Matter* III, p. 50). But there are

no cogent reasons to assume, for example, that more intense motion takes place in the interior of atoms themselves. Occasionally it is strongly emphasized[8] that, according to research by L. Southerns to a fractional accuracy of $5 \cdot 10^{-6}$, the quotient of the two masses has the same value for radioactive uranium oxide as for lead oxide. To be sure, this fact would be of great importance if it were known that in the interior of radioactive atoms intense motions already prevail, such as those exhibited by the emitted α- and β-particles upon explosion. If this were the case, one could not be satisfied with the proposition of proportionality of the two masses as just derived, one would have to demand that it should also be valid for material bodies with intense hidden motion, for example, in such a way that $\nu + 4\alpha = 0$. But to me the hypothesis of violent inner motions in radioactive atoms seems unlikely, especially because it would be hard to understand why it did not produce any radiation of electromagnetic waves. At any rate what is simpler is the notion that also in the interior of radioactive atoms, generally only motions which are to be called very slow compared to the speed of light prevail, but that in this process an atom occasionally reaches an unstable equilibrium state and explodes, and that now its fragments gain the enormous speeds with which they fly apart. But then the result of L. Southerns is explained without further ado.

If one assumes that the hidden motions of the molecules, the atoms, and the ele-
[263] *mentary particles in the interior of the atoms that constitute a material body | are very slow compared to the speed of light, then the principle of the relativity of the gravitational potential yields the law of proportionality of the two masses as an approximate theorem of great accuracy.*

Should the presence of very rapid motions in the interior of atoms really be proved at some time, then there would still be time to examine the theory, as to whether and how it correctly reproduces the action of gravity on these atoms.

Equation (22) allows us to make certain statements about the value of the universal constant Ω. Namely, the ratio

$$\frac{m_g}{m} = \kappa$$

can be specified once one has fixed some system of units (cf. *Theory of Matter* III, p. 42, where instead of κ I used the letter γ). Choosing the erg as unit of energy and mass, the centimeter as unit of length, $1/3 \cdot 10^{10}$ seconds as unit of time, so that the speed of light equals 1, and choosing further the units of the gravitational field such that $\mathfrak{k} = \mathfrak{g}$ in an ideal vacuum, results in:

$$\kappa = 1.016 \cdot 10^{-24},$$

and therefore

$$\Omega = -\frac{(\nu + 3\alpha)}{\kappa} = -(\nu + 3\alpha) \cdot 0.985 \cdot 10^{24}. \tag{23}$$

8 M. Abraham, *Jahrb. d. Radioakt. u. Elektronik* 11, 470 (1915).

Unless $(\nu + 3\alpha)$ happens to be very small, Ω is very large, and the only relevant fact is that it is very large also compared to all changes that the gravitational potential may experience due to the vicinity of large gravitational masses.

For example, let ω_E be the gravitational potential at the surface of the Earth, and let M_g be the gravitational mass and R the radius of the Earth, then:

$$\omega_E - \Omega = \frac{M_g}{4\pi \cdot R} \quad .$$

Substitution for these values in our chosen units

$$M_g = 1.016 \cdot 10^{-24} \cdot 9 \cdot 10^{20} \cdot 5.95 \cdot 10^{27} = 5.44 \cdot 10^{24},$$
$$R = 6.37 \cdot 10^{8} \text{ cm},$$

yields:

$$\omega_E - \Omega = 6.84 \cdot 10^{14}.$$

| So if $(\nu + 3\alpha)$ is of the order of magnitude 1, then $\omega_E - \Omega$ is of the order [264]
$10^{-9} \cdot \Omega$. Likewise the potential ω_s on the surface of the Sun yields

$$\omega_s - \Omega = 2.24 \cdot 10^{18},$$

that is, about $10^{-6}\Omega$ if $(\nu + 3\alpha)$ is of the order of magnitude 1.

Let the potential on the surface of any celestial object be Ω_1, then the quantity a with which the transformation from (4) to (9) is to be executed, is:

$$a = 1 + \frac{\Omega_1 - \Omega}{\Omega} = 1 - \frac{\kappa}{(\nu + 3\alpha)} \cdot (\Omega_1 - \Omega), \tag{24}$$

so a deviates from 1 only by a very small amount. The change of the distances and times under the influence of the changed potential occurs in the ratio:

$$a^{\alpha} = 1 - \frac{\alpha}{\nu + 3\alpha} \cdot \kappa \cdot (\Omega_1 - \Omega), \tag{25}$$

and the change of the density of energy or of the inertial mass in the ratio:

$$a^{\nu} = 1 - \frac{\nu}{\nu + 3\alpha} \cdot \kappa \cdot (\Omega_1 - \Omega). \tag{26}$$

The total inertial mass of a material body changes by the ratio:

$$a^{\nu + 3\alpha} = 1 - \kappa \cdot (\Omega_1 - \Omega), \tag{27}$$

and its total gravitational mass by the ratio:

$$a^{\nu+3\alpha-1} = 1-\left(1-\frac{1}{\nu+3\alpha}\right)\cdot\kappa\cdot(\Omega_1-\Omega). \tag{28}$$

Thus all units change by very small amounts "of the first order" under the influence of the gravitational potential.

6. The theorem of the proportionality of the two masses can be derived in a more intuitive way from the principle of the relativity of the gravitational potential as follows.

Let a material body be located in a region where a gravitational potential ω_∞ exists. By changing masses of material at large distances from the body let the potential be brought to $\omega_\infty+\Delta\omega$, and let this change of potential occur uniformly during a time Δt, so that during Δt the constant gravitational state $\gamma = -\partial\omega/\partial t = -\Delta\omega/\Delta t$ exists at the place considered. I will denote by S a surface that surrounds the body,
[265] but at a sufficient distance from its molecules that | on it the superposition principle is valid for the fields, and so that at points on S the value of the gravitational field caused by the body is constant in time, uninfluenced by the hidden motions of its elementary particles. The surface integral of the field excitation $\mathfrak{k}$ over the surface S then yields the gravitational mass of the body m_g:

$$-\int \mathfrak{k}\cdot d\mathfrak{S} = m_g.$$

Further, during the time Δt and through every element $d\mathfrak{S}$ of the surface there flows a constant energy current of density $\gamma\cdot\mathfrak{k}$ (Theory of Matter III, p. 29). Before and after this time, $\gamma = 0$ on the surface, and therefore also no energy enters or leaves. Thus, as ω_∞ is changed to $\omega_\infty+\Delta\omega$, the body gains the net amount of energy:

$$\begin{aligned} \Delta m &= -\Delta t\cdot\int\gamma\cdot\mathfrak{k}\cdot d\mathfrak{S} = -\gamma\cdot\Delta t\int\mathfrak{k}\cdot d\mathfrak{S}, \\ \Delta m &= -\Delta\omega\cdot m_g. \end{aligned} \tag{29}$$

This is the energy change that was discussed in detail in the introduction as the cause for the attractive effect of gravity in spite of the positive field energy. If we put $\omega_\infty = a\cdot\Omega$ and $\omega_\infty+\Delta\omega = (a+\Delta a)\cdot\Omega$, we can also write equation (29) as follows:

$$\Delta m = -\Omega\cdot m_g\cdot\Delta a,$$

or

$$\Delta m = -\omega_\infty\cdot m_g\cdot\frac{\Delta a}{a}\ . \tag{30}$$

If the principle of the relativity of the gravitational potential is valid, then Δm can be calculated in yet another way. If we denote the density of energy in an arbitrary element of volume dV of the body by W, then:

$$m = \int W \cdot dV, \tag{31}$$

where W is a homogeneous function of degree ν of the variables given on p. 257. If the body were now to experience no change due to the change of the gravitational potential other than the change in measurement units, then during the time Δt one would have to regard the expression (31) as a function of a single variable a, and the change in m would be:

$$\Delta_a m = \left(\int \frac{\partial W}{\partial a} \cdot dV \cdot \int W \cdot \frac{\partial dV}{\partial a} \right) \cdot \Delta a$$

$$\Delta_a m = (\nu + 3\alpha) \cdot \int W \cdot dV \cdot \frac{\Delta a}{a}$$

$$\Delta_a m = (\nu + 3\alpha) \cdot m \cdot \frac{\Delta a}{a} \quad . \tag{32}$$

| If the elementary particles of the body remain at rest in their equilibrium positions, then the supposition just made is certainly satisfied. For the occurrence of the quantity γ during the time Δt does not cause any motion, so the elementary particles are still at rest in the equilibrium positions after Δt has passed, and the body has experienced no other changes than those subject to the changes of a. In this case we can put $\Delta_m = \Delta_a m$, and the combination of equation (31) with (32) yields: [266]

$$-\omega_\infty \cdot m_g = (\nu + 3\alpha) \cdot m.$$

Thus for a body with motionless elementary particles we have found equation (21) in a second way.

However, if the elementary particles execute hidden motions, then it is to be expected that the occurrence of the quantity γ during the time Δt influences these motions, particularly because the superposition principle is not valid in the interior of matter, and because therefore all state variables are influenced by the value of γ. After the time Δt has passed, the average value of the hidden motion of the elementary particles will therefore have become different than before Δt due to the action of γ. We can express this intuitively by saying that a change in the gravitational potential is to be associated with a small adiabatic temperature change of the body. That is, not only is there a change in the temperature as such, which is a measure of the random motion of the molecules, but there is also a change in that quantity which we may characterize as the temperature of motion of the elementary particles inside the atom. The temperature inside the atom does not have to be associated with the body temperature proper. I will denote by ΔQ the small amount of energy that one would have to transfer to the body in order to reduce the temperature of the hidden motion of its elementary particles to the values they had before Δt, so that the net change in the body's energy would be $\Delta_a m$. This is the latent or bound energy gained by the

body during an "isothermal" change of ω. Its energy change upon an "adiabatic" change is therefore:

$$\Delta m = \Delta_a m - \Delta Q. \tag{33}$$

Now I denote by $(\partial Q/\partial \omega)_{is}$ the increase of the latent energy of the body during an "isothermal" change of the gravitational potential in comparison with the increase in the potential, then we have:

$$\Delta Q = \omega_\infty \cdot \left(\frac{\partial Q}{\partial \omega}\right)_{is} \cdot \frac{\Delta a}{a}. \tag{34}$$

[267] | If we substitute the expressions of equation (30), (32), (34) into equation (33) then the result is:

$$(\nu + 3\alpha) \cdot m - \omega_\infty \cdot \left(\frac{\partial Q}{\partial \omega}\right)_{is} = -\omega_\infty \cdot m_g. \tag{35}$$

The statement on proportionality of the two masses is approximately valid if the "latent" energy ΔQ is vanishingly small in comparison with the "free" energy Δm.

Comparing equation (19), found from Laue's theorem, with (35) we find:

$$(\nu + 4\alpha) \cdot \int \overline{\gamma \cdot \chi} \cdot dV = \omega_\infty \cdot \left(\frac{\partial Q}{\partial \omega}\right)_{is}. \tag{36}$$

RESULTS AND PROSPECTS

1. The principle of the relativity of the gravitational potential is established as the simplest expression of the fact that the gravitational potential in general has no perceptible influence on material processes, although it occurs as an independent quantity of state in the basic equations of the physics of the aether. We succeeded in formulating the principle in a quite general fashion, without making special assumptions about the form in which the gravitational potential enters into the basic equations of aether physics.

2. From the principle of the relativity of the gravitational potential one can derive theoretically the well-known empirical law of the proportionality of the gravitational and inertial mass of all material bodies. It is true that this law may have only approximate validity for bodies whose elementary particles execute random hidden motions. But if the speeds of the hidden motions are very small compared to the speed of light, the accuracy to which the law is valid can be so great that one cannot find deviations experimentally. If one wanted the law to be valid in general and with mathematical precision, one would have to supplement the principle of the relativity of the gravitational potential with an extra assumption.

3. The additional term that implies deviation from the mathematically exact validity of the proportionality of the two masses can be given an interesting interpretation. If the gravitational potential experiences a change at the place where a material body

is located, maybe due to displacement of a distant large and heavy mass, then simultaneously with this there is a change not only of the body's energy | content, but also [268]
in general of its temperature of the molecular random motion as well as of the interatomic motion. If the body should change strictly isothermally with the gravitational potential, so that the temperatures of the hidden motions of its elementary particles all remain constant, then the change of its free energy due to the potential change must be supplemented by a a change of its latent energy, for example by radiation. This supplied or removed latent energy provides the measure of our additional term. If it is very small compared to the change of the free energy, the deviation from proportionality of the two masses is small; if this latent energy vanishes, the theorem of the proportionality is mathematically exact.

4. As the next goal of the theory of matter, the task of setting up a principle for the electromagnetic four potential that is analogous to the principle of the relativity of the gravitational potential, and thereby providing an explanation for the lack of perceptible influence of the four potential on material processes presents itself.

MAX BORN

THE MOMENTUM-ENERGY LAW IN THE ELECTRODYNAMICS OF GUSTAV MIE

Originally published as "Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie" in Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen 1914, 1, pp. 23–36. Submitted by Mr. Hilbert during the meeting on 20th December 1913.

INTRODUCTION: THE MATHEMATICAL FORM OF MIE'S ELECTRODYNAMIC CONCEPTION OF THE WORLD

Whereas the electron theory developed by H. A. Lorentz requires certain hypotheses about the structure of the electron (e.g. the hypothesis regarding the rigidity in the usual sense, or in the context of the theory of relativity), Gustav Mie[1] set himself the task of trying to modify Maxwell's equations in such a way that the existence of electrons ("nodes" of the field) and, even more generally, the existence of material atoms and molecules follows necessarily from the new equations. The fact that without the addition of new forces, stable accumulations of charge, as represented by electrons, are incompatible with the usual differential equations of the magnetic field is closely linked to the linearity of these equations. Therefore, it was first of all necessary to relinquish linearity. Mie carried out this idea in the most general and elegant manner which can be imagined in the framework of today's physics borne from Lagrange's analytical mechanics. To illustrate the type of generalization of the fundamental equations, it is perhaps | best to start with the equation of motion of a system of [24]
masses with one degree of freedom q. If

$$\Phi(\dot{q}, q) = T - U$$

represents the Lagrangian (difference between kinetic and potential energy), then it is well known that from the variation of the Hamiltonian integral

1 G. Mie, *Grundlagen einer Theorie der Materie*. 3rd. communication in *Ann. d. Phys.* (4), vol. 37, p. 511; vol. 39, p. 1; vol. 40, p. 1.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

$$\int_{t_1}^{t_2} \Phi(\dot{q}, q)\, dt \tag{1}$$

one obtains the equations of motion in the form

$$\frac{d}{dt}\frac{\partial \Phi}{\partial \dot{q}} - \frac{\partial \Phi}{\partial q} = 0. \tag{2}$$

The transition from the usual equations of the electromagnetic field to Mie's fundamental equations can then be considered to be parallel to the transition from a quasi-elastic system where Φ has the form $\Phi = \frac{a}{2}\dot{q}^2 + \frac{b}{2}q^2$ to a system where Φ is a completely arbitrary function of $\dot{q}$ and q. In this process, the form of the differential equation (2) remains completely preserved. Indeed, in the final analysis, Mie's theory aims to show that the field equations of electron theory are variational derivatives of a variational principle completely analogous to (1), except that there are 4 functions of 4 variables, where Φ is a certain quadratic form of the field quantities, and that then, as in the mechanical example shown above, the form of the fundamental equations remains completely preserved if Φ becomes an arbitrary function of the field quantities. Therefore, one can say that the equations of Mie achieve the same for electrodynamics as the Lagrange equations of the second kind achieve for the mechanics of systems of point particles [*Punktsysteme*]. They offer a formal scheme which, through an appropriate choice of the function Φ, can be adjusted to the specific properties of the system. As the aim of the mechanistic explanation of nature in the past was to derive a Lagrangian function Φ for the interaction of atoms and to derive all physical and chemical properties of matter, so Mie now sets for himself the task of selecting his "world-function" Φ in such a way that on the basis of its differential equations the existence of the electron and the atoms, as well as the totality of their interactions follows. I would like to view this requirement of Mie as the mathematical
[25] content of that | program which considers the aim of physics to be the construction of an "electromagnetic worldview."

In the following, I would like to make a contribution to the clarification of the mathematical structure of Mie's fundamental equations. The variational problem of Mie is still not the most general one can devise for the four-dimensional continuum, and one is well advised to compare it with the most general, in order to determine what are the properties which have to be attributed to the four-dimensional continuum (the aether), in order to obtain specifically Mie's laws. It will turn out that these are *not* the properties of an *elastic* body. The four-dimensional theory of elasticity compatible with the principle of relativity has been exhaustively treated by Herglotz[2] and is obtained through a different specialization of our variational principle. Mie's four-dimensional continuum corresponds rather to the three-dimensional aether of MacCullagh,[3] who, from the assumption that the vortices of the aether and not its

2 *Ann. d. Phys.* (4), vol. 36, p. 493.

deformations store energy, obtains equations identical with Maxwell's equations for stationary electrodynamic processes. The analogy of Mie's theory with Lagrangian mechanics is manifested most clearly by considering the law of conservation of energy. It is known that for a variational problem of the form (1), there always exists an integral of the differential equations (2), expressing conservation of energy, if the independent variable t does not appear explicitly in Φ (t is then a "cyclic coordinate"). Because then one has

$$\frac{d\Phi}{dt} = \frac{\partial\Phi}{\partial\dot{q}}\ddot{q} + \frac{\partial\Phi}{\partial q}\dot{q},$$

and if one adds to this equation (2) multiplied by $\dot{q}$, one obtains

$$\frac{d\Phi}{dt} = \frac{d}{dt}\left(\dot{q}\frac{\partial\Phi}{\partial\dot{q}}\right); \tag{3}$$

If one introduces as "energy" the Legendre transform of Φ:

$$W = \Phi - \dot{q}\frac{\partial\Phi}{\partial\dot{q}},$$

then one can write equation (3) in the form

$$\frac{dW}{dt} = 0 \quad \text{or} \quad W = \text{const.} \tag{3'}$$

which represents the law of conservation of energy. [26]

In Mie's electrodynamics there also exists a momentum-energy conservation law which plays a significant role in all the new dynamical theories based on the principle of relativity. The law consists of 4 equations, the first three express the conservation of momentum, the last the conservation of energy. Mie obtains the last of these equations by calculation and the others on the basis of symmetry requirements demanded by the principle of relativity. I will show in the following that these 4 equations are precise generalizations of equation (3) for the case of 4 variables. The requirement for them to be valid is, as before, that the function Φ does not contain the 4 independent variables explicitly, and the proof follows along the same lines of reasoning we used when deriving equation (3). In this process, the structure of Mie's formulae for the energy quantities will emerge which, at first sight, is not readily apparent.

3 *Irish. Trans.* 21.

1. THE VARIATIONAL PRINCIPLE OF STATICS FOR A FOUR-DIMENSIONAL CONTINUUM

One will be able to describe the deformation of a four-dimensional continuum by giving the projections u_1, u_2, u_3, u_4, of the deformations of its points with respect to 4 orthogonal axes as functions of the coordinates x_1, x_2, x_3, x_4:

$$u_\alpha = u_\alpha(x_1, x_2, x_3, x_4), \qquad \alpha = 1, \ldots 4. \tag{4}$$

We further use the abbreviation

$$\frac{\partial u_\alpha}{\partial x_\beta} = a_{\alpha\beta}. \tag{5}$$

All properties of the continuum should now be determined through the function Φ of the displacements u_α and their derivatives $a_{\alpha\beta}$, and the resulting deformations should be determined by the requirement that variations of the four-dimensional integral over the four-dimensional space

$$\int \Phi(a_{11}, a_{12}, a_{13}, a_{14}; a_{21}, \ldots a_{44}; u_1, \ldots u_4) dx_1 dx_2 dx_3 dx_4 \tag{6}$$

vanish. |

[27] If we now use the abbreviation[4]

$$\frac{\partial \Phi}{\partial a_{\alpha\beta}} = X_{\alpha\beta}, \qquad \frac{\partial \Phi}{\partial u_\alpha} = X_\alpha, \tag{7}$$

then this requirement yields the 4 differential equations:

$$\sum_\gamma \frac{\partial X_{\beta\gamma}}{\partial x_\gamma} - X_\beta = 0, \tag{8}$$

which express the requirement for equilibrium and correspond to equation (2) in the introduction.

2. FIRST SPECIAL CASE OF THE PRINCIPLE: HERGLOTZ' THEORY OF ELASTICITY

In the theory of relativity, x_1, x_2, x_3, represent the space coordinates and x_4 is the time multiplied by the imaginary unit i and the speed of light. The statics of the four-dimensional continuum is then nothing other than the dynamics of the three-dimensional one.

Therefore, the theory of elasticity, which has been adapted by Herglotz to satisfy the principle of relativity, must appear as a special case of our principle (6).

---

4 In the following all indices shall run through the values 1, 2, 3, 4, and all sums should extend over these values.

I will briefly outline how the quantities appearing in this process are to be interpreted and how the function Φ must be specified. The independent variables x_1, x_2, x_3, have to be considered as parameters ξ, η, ζ, which at a given instant fix the position of the points of the body; x_4 is set to $ic\tau$, where τ is a "time-like" parameter which otherwise is totally arbitrary. u_1, u_2, u_3, are the coordinates x, y, z, of the points of the body at an arbitrary time $t = u_4/ic$. Then, the quantities $a_{\alpha\beta}$, for $\alpha, \beta = 1, 2, 3$ are obviously determined by the strain in the body, whereas $a_{14}/a_{44}, a_{24}/a_{44}, a_{34}/a_{44}$ are the velocity components. The function Φ is now specified through the requirement that the integral (6) neither changes its value under a Lorentz transformation of the variables u_1, u_2, u_3, u_4 (rotation of the four-dimensional space) nor under a change of the time | parameter τ. Consequently, Φ is not a [28] function of all 16 quantities $a_{\alpha\beta}$, but depends only on a combination of 6 of them, the "rest-deformations" $e_{11}, e_{22}, e_{33}, e_{23}, e_{31}, e_{12}$. These quantities, which I introduced first,[5] are a measure of the deformation of the volume element as measured by a co-moving observer. Most remarkable in this formulation is the absence of the kinetic energy. Instead, the velocities appear in the rest-deformations $e_{\alpha\beta}$. Herglotz extensively examined the laws of motion that arise from the interpretation of these quantities, equation (8), and showed that the ordinary mechanics of elastic bodies is a limiting case of this theory.

3. SECOND SPECIAL CASE OF THE PRINCIPLE: MIE'S ELECTRODYNAMICS

The theory of Mie is quite a different special case of the variational principle (6). Before we interpret the electrodynamic significance of the quantities, we want to present the characteristic specification of the function Φ, which shapes the entire theory: Φ shall only be a function of the differences

$$a_{\alpha\beta} - a_{\beta\alpha} = \frac{\partial u_\alpha}{\partial x_\beta} - \frac{\partial u_\beta}{\partial x_\alpha} \tag{9}$$

This formulation applied to three-dimensional space leads exactly to the theory of MacCullagh mentioned in the introduction. Therefore, one can interpret the formulae here in the same manner as is done there. The quantities (9) are namely the components of the infinitesimal rotation of the volume elements of the continuum, the "rotation components." In the theory of MacCullagh, the energy of the aether depends only on these rotations, but not on the deformation of the aether. It is clear that we can conceptualize Mie's theory in the same way if instead of aether we say "four-dimensional world." We leave it open whether a mechanistic interpretation in the usual sense of this formulation is possible and we restrict ourself to the assertion that it contains the entire electrodynamics of Mie (and, as a special case, also the classical electron theory).

---

5 *Ann. d. Phys.* (4), vol. 30, 1909, p. 1.

Let us now turn to the physical interpretation and description of the quantities
[29] appearing here. In Mie's theory, x_1, x_2, x_3, x_4 are | nothing but the coordinates and the time x, y, z, ict. Furthermore, Mie writes for

$$\begin{aligned} &u_1, u_2, u_3, u_4 \\ &f_x, f_y, f_z, i\varphi. \end{aligned} \tag{10}$$

These primary quantities[6] characterizing the aether correspond to the components of the four-potential in the theory of electrons.

The components of rotation (9) appear in Mie's theory as components of the 6-vector $(\mathfrak{b}, -i\mathfrak{e})$, where $\mathfrak{b}$ represents the magnetic induction and $\mathfrak{e}$ the electric field strength according to the scheme

$$(a_{\alpha\beta} - a_{\beta\alpha}) = \begin{vmatrix} 0 & -\mathfrak{b}_z & \mathfrak{b}_y & i\mathfrak{e}_x \\ \mathfrak{b}_z & 0 & -\mathfrak{b}_x & i\mathfrak{e}_y \\ -\mathfrak{b}_y & \mathfrak{b}_x & 0 & i\mathfrak{e}_z \\ -i\mathfrak{e}_x & -i\mathfrak{e}_y & -i\mathfrak{e}_z & 0 \end{vmatrix}. \tag{11}$$

This can also be written as

$$(\mathfrak{b}, -i\mathfrak{e}) = \mathfrak{Curl}(f, i\varphi) \tag{11'}$$

or as

$$\mathfrak{b} = \operatorname{curl} f, \qquad \mathfrak{e} = -\operatorname{grad}\varphi - \frac{\partial f}{\partial t}. \tag{11''}$$

With these symbols, Φ is seen to be a function of the components of the vectors $\mathfrak{b}, \mathfrak{e}, f$, and of the scalar φ:

$$\Phi(\mathfrak{b}_x, \mathfrak{b}_y, \mathfrak{b}_z, \ \mathfrak{e}_x, \mathfrak{e}_y, \mathfrak{e}_z; f_x, f_y, f_z, \varphi), \tag{6'}$$

where the fundamental assumption, that Φ depends only on the rotations $a_{\alpha\beta} - a_{\beta\alpha}$, is manifested. But at the same time, this also implies that the vectors $\mathfrak{e}, \mathfrak{b}$ satisfy the
[30] one quadruple of | Maxwell's equations, namely:

$$\operatorname{Div}(\mathfrak{b}, -i\mathfrak{e}) = 0 \tag{12}$$

---

6 In this presentation, f_x, f_y, f_z, φ, should be designated as "extensive quantities" [*Quantitätsgrößen*] since they have the character of displacement components of the four-dimensional continuum. The $\mathfrak{v}_x, \mathfrak{v}_y, \mathfrak{v}_z$, to be defined momentarily, would then be introduced as "intensive quantities" [*Intensitätsgrößen*]. That Mie proceeds here, as with the division of field-vectors into extensive and intensive quantities, in exactly the opposite way has its origin in the fact that the formulation of his expressions is closer to the physical conceptualization of electric density, displacement, field strength etc. In addition, Mie uses a different variational principle, which arises from ours via a Legendre transformation, and which readily suggests his choice of division of the quantities. Since Mie's variational principle requires additional conditions which cannot be readily incorporated into the formulation of the statics of the four-dimensional continuum, I have preferred the approach presented here.

or

$$\operatorname{curl}\mathfrak{e} + \frac{\partial \mathfrak{b}}{\partial t} = 0, \qquad \operatorname{div}\mathfrak{b} = 0; \tag{12'}$$

since these equations follow directly from (11′) and (11″) respectively. The differential equations (8), however, are nothing but the second quadruple of Maxwell's equations. To show this we set, like Mie:[1]

$$\begin{aligned}
&\frac{\partial \Phi}{\partial \mathfrak{b}_x} = \mathfrak{h}_x, \qquad \frac{\partial \Phi}{\partial \mathfrak{e}_x} = -\mathfrak{d}_x, \\
&\frac{\partial \Phi}{\partial \mathfrak{b}_y} = \mathfrak{h}_y, \qquad \frac{\partial \Phi}{\partial \mathfrak{e}_y} = -\mathfrak{d}_y, \\
&\frac{\partial \Phi}{\partial \mathfrak{b}_z} = \mathfrak{h}_z, \qquad \frac{\partial \Phi}{\partial \mathfrak{e}_z} = -\mathfrak{d}_z, \\
&\frac{\partial \Phi}{\partial f_x} = -\mathfrak{v}_x, \qquad \frac{\partial \Phi}{\partial f_y} = -\mathfrak{v}_y, \qquad \frac{\partial \Phi}{\partial f_z} = -\mathfrak{v}_z, \qquad \frac{\partial \Phi}{\partial \varphi} = \rho.
\end{aligned} \tag{13}$$

Then, the 4 quantities X_α of the general theory become identified with the quantities $-\mathfrak{v}_x, -\mathfrak{v}_y, -\mathfrak{v}_z, -i\rho$, and the 16 quantities $X_{\alpha\beta}$ with the components of the vectors $\mathfrak{h}$ and $\mathfrak{d}$ as illustrated in the matrix equation

$$(X_{\alpha\beta}) = \begin{vmatrix} 0 & -\mathfrak{h}_z & \mathfrak{h}_y & i\mathfrak{d}_x \\ \mathfrak{h}_z & 0 & -\mathfrak{h}_x & i\mathfrak{d}_y \\ -\mathfrak{h}_y & \mathfrak{h}_x & 0 & i\mathfrak{d}_z \\ -i\mathfrak{d}_x & -i\mathfrak{d}_y & -i\mathfrak{d}_z & 0 \end{vmatrix} \tag{14}$$

With this notation the equations (8) turn into

$$\operatorname{Div}(\mathfrak{h}, -i\mathfrak{d}) = (\mathfrak{v}, i\rho) \tag{15}$$

or

$$\operatorname{curl}\mathfrak{h} - \frac{\partial \mathfrak{d}}{\partial t} = \mathfrak{v}, \qquad \operatorname{div}\mathfrak{d} = \rho. \tag{15'}$$

From these, one recognizes that ρ is the electric charge density, $\mathfrak{v}$ the convection current (charge times velocity), $\mathfrak{h}$ the magnetic field strength and $\mathfrak{d}$ the electric displacement. We also see that these quantities, according to equation (14), in terms of the picture of the | statics of the four-dimensional continuum, correspond to stresses [31] and forces.

The equation of continuity for the electric current follows from equation (15)

$$\operatorname{Div}(\mathfrak{v}, i\rho) = 0 \tag{16}$$

or

$$\operatorname{div}\mathfrak{v} + \frac{\partial \rho}{\partial t} = 0. \tag{16'}$$

However, Φ is still an arbitrary function of its 10 arguments. We see that Maxwell's equations (12) and (15) are formally valid for any function Φ.

However, if one wants to maintain the validity of the principle of relativity, the choice of Φ has to be restricted. Obviously, Φ is then not allowed to depend explicitly on all of the 10 arguments, but only on such combinations of them as are invariant under Lorentz transformations. Mie has shown that there exist four such invariants which are independent of one another. In our representation we could for instance choose the following 4 invariants:

1. The length of the four-vectors $(f, i\varphi)$:

$$\chi = \sqrt{\varphi^2 - f^2},$$

2. The absolute magnitude of the six-vector $(\mathfrak{b}, -i\mathfrak{e})$:

$$\eta = \sqrt{\mathfrak{e}^2 - \mathfrak{b}^2},$$

3. The scalar product of the six-vector $(\mathfrak{b}, -i\mathfrak{e})$ with its dual vector $(-i\mathfrak{e}, \mathfrak{b})$:

$$\kappa = (\mathfrak{b}\mathfrak{e})$$

4. As the simultaneous invariant of the four-vector and of the six-vector one can take the square of the length of the four-vector obtained from the multiplication of the two original vectors:

$$\lambda^2 = ([f\mathfrak{b}] + \varphi\mathfrak{e})^2 - (f\mathfrak{e})^2.$$

[32] Φ can still be chosen as an arbitrary function of these 4 | arguments.

The aim of physical research then, as suggested by the theory of Mie, is to account, through an appropriate choice of the function $\Phi(\chi, \eta, \kappa, \lambda)$, for all the electromagnetic properties[7] of electrons and atoms.

In this, we have exactly the continuation of Lagrange's magnificent program.

The classical theory of electrons is formally a special case of Mie's theory, but not in the strict sense. Indeed one obtains its field equations by simply setting:

$$\Phi = \frac{1}{2}(\mathfrak{b}^2 - \mathfrak{e}^2) - (f\mathfrak{v}) + \varphi\rho, \tag{17}$$

where $\mathfrak{v}_x, \mathfrak{v}_y, \mathfrak{v}_z$, and ρ are considered to be *given* functions of space and time which describe the motion of the electrons.

But then Φ is no longer a function of only the 4 invariants $\chi, \eta, \kappa, \lambda$, but in addition depends explicitly on x, y, z, t, which however, is excluded in Mie's theory on

7 We are excluding gravitation here.

principle. In Mie's theory, the forces that hold electrons and atoms together should arise naturally from the formulation of Φ, whereas in the classical theory of electrons the forces have to be specifically added.

4. THE MOMENTUM-ENERGY LAW FOR THE GENERAL CASE OF THE FOUR-DIMENSIONAL CONTINUUM

The assumption of Mie just emphasized, that the function Φ is independent of x, y, z, t, is also the real mathematical reason for the validity of the momentum-energy-law.

In order to show that, we first consider, as in section 1, a general four-dimensional continuum whose equilibrium is determined by equation (8). We assert that for these differential equations, a law, analogous to the energy law (3′) of Lagrangian mechanics, is always valid as soon as one of the 4 coordinates x_α does not appear explicitly in Φ.

Then one obtains by differentiation of Φ with respect to x_α:

$$\frac{\partial \Phi}{\partial x_\alpha} = \sum_{\beta,\gamma} X_{\beta\gamma} \frac{\partial^2 u_\beta}{\partial x_\alpha \partial x_\gamma} + \sum_\beta X_\beta \frac{\partial u_\beta}{\partial x_\alpha},$$

and if one now adds equations (8) multiplied by the quantities $\frac{\partial u}{\partial x_\alpha}$ to the above equation one obtains: |

$$\frac{\partial \Phi}{\partial x_\alpha} = \sum_\gamma \frac{\partial}{\partial x_\gamma} \left(\sum_\beta X_{\beta\gamma} a_{\beta\alpha} \right). \tag{18}$$ [33]

This is the formula corresponding to the energy conservation law in mechanics. If Φ is independent of all 4 coordinates x_α, then (18) is valid for $\alpha = 1, 2, 3, 4$. These 4 equations are to be designated the momentum-energy theorem. They can also be summarized by the symbolic equation

$$\mathrm{Div} T = 0, \tag{18'}$$

if the 16 components $T_{\alpha\beta}$ of the matrix T are defined as

$$T_{\alpha\beta} = \Phi \delta_{\alpha\beta} - \sum_\gamma a_{\gamma\alpha} X_{\gamma\beta} \tag{19}$$

where

$$\delta_{\alpha\beta} = \begin{cases} 1 & \text{for } \alpha = \beta \\ 0 & \text{for } \alpha \neq \beta \end{cases}.$$

In the matrix calculus, (19) can be written as:

$$T = \Phi - \bar{a} X, \tag{19'}$$

where $\bar{a} = (a_{\beta\alpha})$ is the transpose of the matrix $a = (a_{\alpha\beta})$.[8]

5. SPECIAL FORM OF THE MOMENTUM-ENERGY EQUATION FOR THE CASE OF MIE'S ELECTRODYNAMICS.

The equations described by Mie as the momentum-energy equations are essentially nothing but the general equations (18) and (18′) respectively. A minor mathematical transformation leads to the formulae of Mie. In order to see why the transformation is necessary, it is best to consider the stress-energy tensor in the succinct symbolic form (19′). Keeping in mind the electrodynamic significance of the quantities $a_{\alpha\beta}$ and $X_{\alpha\beta}$ (equations (11) and (14)), we see that although the quantities $X_{\alpha\beta}$ can be expressed directly through the components of the field vectors, the $a_{\alpha\beta}$ cannot; rather, only the combinations $a_{\alpha\beta} - a_{\beta\alpha}$ whose matrix is to be denoted by $a - \bar{a}$ have a physical meaning. Therefore, we will have to transform equation (19′) in such a manner that it contains the difference-matrix $a - \bar{a}$. |

[34] Naturally, we may not simply add aX to T, because then (18′) would cease to be valid. Nevertheless, one can try to define a matrix ψ in such a way that the divergence equations (18′) remains valid for the matrix

$$S = \Phi + (a - \bar{a})X + \psi. \tag{20}$$

If we denote the added matrix $\psi + aX$ by ω, so that $S = T + \omega$, then we also require that

$$\mathrm{Div}\omega = 0. \tag{21}$$

We now show that (21) is satisfied by the matrix

$$\psi_{\alpha\beta} = -u_\alpha X_\beta \tag{22}$$

provided that the matrix X is skew-symmetric:

$$X_{\alpha\beta} = -X_{\beta\alpha}, \quad \text{or} \quad X = -\bar{X}. \tag{23}$$

Then, because of (8), we have

$$\begin{aligned}\omega_{\alpha\beta} &= \sum_\gamma a_{\alpha\gamma}X_{\gamma\beta} - u_\alpha X_\beta \\ &= \sum_\gamma \left(\frac{\partial u_\alpha}{\partial x_\gamma}X_{\gamma\beta} + u_\alpha \frac{\partial X_{\gamma\beta}}{\partial x_\gamma}\right) \\ &= \sum_\gamma \frac{\partial}{\partial x_\gamma}(u_\alpha X_{\alpha\beta})\end{aligned}$$

8 The product of two matrices is that matrix whose element with the subscripts α, β arises from multiplying the row α by the column β.

and by using (23) again

$$\sum_{\beta} \frac{\partial \omega_{\alpha\beta}}{\partial x_\beta} = \sum_{\beta\gamma} \frac{\partial^2 u_\alpha X_{\gamma\beta}}{\partial x_\beta \partial x_\gamma} = -\sum_{\beta\gamma} \frac{\partial^2 u_\alpha X_{\gamma\beta}}{\partial x_\beta \partial x_\gamma} = 0;$$

i.e. equation (21) is satisfied.

As a glance at the scheme (14) shows, the condition (23) is met in Mie's theory precisely because of the requirement that Φ is only a function of the differences $a_{\alpha\beta} - a_{\beta\alpha}$. Hence, one can write the energy-momentum equation in the form

$$\mathrm{Div} S = 0, \tag{24}$$

where S is defined by (20) and (22).

The mathematical structure of the law is especially transparent in this form. |

If we introduce the electromagnetic notation, then the matrix equation (20) becomes [35]

$$S = \begin{vmatrix} \Phi & 0 & 0 & 0 \\ 0 & \Phi & 0 & 0 \\ 0 & 0 & \Phi & 0 \\ 0 & 0 & 0 & \Phi \end{vmatrix} + \begin{vmatrix} 0 & -\mathfrak{b}_z & \mathfrak{b}_y & i\mathfrak{e}_x \\ \mathfrak{b}_z & 0 & -\mathfrak{b}_x & i\mathfrak{e}_y \\ -\mathfrak{b}_z & \mathfrak{b}_x & 0 & i\mathfrak{e}_z \\ -i\mathfrak{e}_x & -i\mathfrak{e}_y & -i\mathfrak{e}_z & 0 \end{vmatrix} \cdot \begin{vmatrix} 0 & \mathfrak{h}_z & \mathfrak{h}_y & i\mathfrak{d}_x \\ \mathfrak{h}_z & 0 & -\mathfrak{h}_x & i\mathfrak{d}_y \\ -\mathfrak{h}_y & \mathfrak{h}_x & 0 & i\mathfrak{d}_z \\ -i\mathfrak{d}_x & -i\mathfrak{d}_y & -i\mathfrak{d}_z & 0 \end{vmatrix} + \begin{vmatrix} f_x\mathfrak{v}_x & f_x\mathfrak{v}_y & f_x\mathfrak{v}_z & if_x\rho \\ f_y\mathfrak{v}_x & f_y\mathfrak{v}_y & f_y\mathfrak{v}_z & if_y\rho \\ f_z\mathfrak{v}_x & f_z\mathfrak{v}_y & f_z\mathfrak{v}_z & if_z\rho \\ i\varphi\mathfrak{v}_x & i\varphi\mathfrak{v}_y & i\varphi\mathfrak{v}_z & -\varphi\rho \end{vmatrix} \tag{20'}$$

or carrying out the multiplication:

$$S = \begin{vmatrix} \Phi - \mathfrak{b}_y\mathfrak{h}_y - \mathfrak{b}_z\mathfrak{h}_z + \mathfrak{e}_x\mathfrak{d}_x + f_x\mathfrak{v}_x, & \mathfrak{e}_x\mathfrak{d}_y + \mathfrak{h}_x\mathfrak{b}_y + f_x\mathfrak{v}_y, \\ \mathfrak{e}_y\mathfrak{d}_x + \mathfrak{h}_y\mathfrak{b}_x + f_y\mathfrak{v}_x, & \Phi - \mathfrak{b}_z\mathfrak{h}_z - \mathfrak{b}_x\mathfrak{h}_x + \mathfrak{e}_y\mathfrak{d}_y + f_y\mathfrak{v}_y, \\ \mathfrak{e}_z\mathfrak{d}_x + \mathfrak{h}_z\mathfrak{b}_x + f_z\mathfrak{v}_x, & \mathfrak{e}_z\mathfrak{d}_y + \mathfrak{h}_z\mathfrak{b}_y + f_z\mathfrak{v}_y, \\ -i(\mathfrak{e}_y\mathfrak{h}_z - \mathfrak{e}_z\mathfrak{h}_y - \varphi\mathfrak{v}_x), & -i(\mathfrak{e}_z\mathfrak{h}_x - \mathfrak{e}_x\mathfrak{h}_z - \varphi\mathfrak{v}_y), \\ \mathfrak{e}_x\mathfrak{d}_z + \mathfrak{h}_x\mathfrak{b}_z + f_x\mathfrak{v}_z, & -i(\mathfrak{d}_y\mathfrak{b}_z - \mathfrak{d}_z\mathfrak{b}_y - \rho f_x), \\ \mathfrak{e}_y\mathfrak{d}_z + \mathfrak{h}_y\mathfrak{b}_z + f_y\mathfrak{v}_z, & -i(\mathfrak{d}_z\mathfrak{b}_x - \mathfrak{d}_x\mathfrak{b}_z - \rho f_y), \\ \Phi - \mathfrak{b}_x\mathfrak{h}_x - \mathfrak{b}_y\mathfrak{h}_y + \mathfrak{e}_z\mathfrak{d}_z + f_z\mathfrak{v}_z, & -i(\mathfrak{d}_x\mathfrak{b}_y - \mathfrak{d}_y\mathfrak{b}_x - \rho f_z), \\ -i(\mathfrak{e}_x\mathfrak{h}_y - \mathfrak{e}_y\mathfrak{h}_x - \varphi\mathfrak{v}_z), & \Phi + \mathfrak{e}_x\mathfrak{d}_x + \mathfrak{e}_y\mathfrak{d}_y + \mathfrak{e}_z\mathfrak{d}_z - \varphi\rho \end{vmatrix} \tag{20''}$$

This is precisely the stress-energy matrix presented by Mie.

Mie then showed that this matrix is symmetric provided Φ depends only on the 4 invariants $\chi, \eta, \kappa, \lambda$. This proof, which is carried out by simple calculation, cannot be significantly simplified by our method of presentation.

It is perhaps not superfluous to emphasize that the energy-momentum equation of the classical electron theory does not arise as a special case by using Φ in $(20'')$ as formulated in (17), because then Φ is not independent of x, y, z, t, since $\mathfrak{v}$ and ρ depend on position and time and thus, our line of proof becomes invalid. One can also easily see, by substituting in (24) for Φ as formulated in (17), that the result is at variance with the energy-momentum law of the classical electron theory. However, if one adds to (24) the terms that arise by differentiating (17) with respect to x, y, z, t, which arise because of their dependence on $\mathfrak{v}$ and ρ, and which cannot be written in
[36] the form of a four-dimensional | divergence, then one obtains the energy-momentum law of the electron theory in its usual form. With respect to the corresponding question in the electrodynamic theory of moving material bodies the same is to be said. None of the available formulations for the stress-energy-matrix, neither Minkowski's unsymmetric one, nor Abraham and Laue's symmetric one, fall directly under Mie's scheme, yet the same method can be employed here as well.

EDITORIAL NOTE

[1] In the last line of eqs. (13), ∂f_x is misprinted in the original as $\partial \mathfrak{b}_x$.

INCLUDING GRAVITATION IN A UNIFIED THEORY OF PHYSICS

LEO CORRY

THE ORIGIN OF HILBERT'S AXIOMATIC METHOD[1]

1. AXIOMATICS, GEOMETRY AND PHYSICS IN HILBERT'S EARLY LECTURES

This chapter examines how Hilbert's axiomatic approach gradually consolidated over the last decade of the nineteenth century. It goes on to explore the way this approach was actually manifest in its earlier implementations.

Although geometry was not Hilbert's main area of interest before 1900, he did teach several courses on this topic back in Königsberg and then in Göttingen. His lecture notes allow an illuminating foray into the development of Hilbert's ideas and they cast light on how his axiomatic views developed.[2]

1.1 Geometry in Königsberg

Hilbert taught projective geometry for the first time in 1891 (Hilbert 1891). What already characterizes Hilbert's presentation of geometry in 1891, and will remain true later on, is his clearly stated conception of this science as a natural one in which, at variance with other mathematical domains, sensorial intuition— *Anschauung*—plays a fundamental role that cannot be relinquished. In the introduction to the course, Hilbert formulated it in the following words:

> Geometry is the science that deals with the properties of space. It differs essentially from pure mathematical domains such as the theory of numbers, algebra, or the theory of functions. The results of the latter are obtained through pure thinking... The situation is completely different in the case of geometry. I can never penetrate the properties of space by pure reflection, much as I can never recognize the basic laws of mechanics, the law of gravitation or any other physical law in this way. Space is not a product of my reflections. Rather, it is given to me through the senses. I thus need my senses in order to fathom its properties. I need intuition and experiment, just as I need them in order to figure out physical laws, where also matter is added as given through the senses.[3]

---

1 This chapter is based on extracts from (Corry 2004), in particular on chapters 2, 3, and 5.

2 An exhaustive analysis of the origins of *Grundlagen der Geometrie* based on these lecture notes and other relevant documents was first published in (Toepell 1986). Here we draw directly from this source.

3 The German original is quoted in (Toepell 1986, 21). Similar testimonies can be found in many other manuscripts of Hilbert's lectures. Cf., e.g., (Toepell 1986, 58).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

The most basic propositions related to this intuition concern the properties of incidence, and in order to express them conveniently it is necessary to introduce "ideal elements." Hilbert stressed that these are to be used here only as a shorthand with no metaphysical connotations.

In the closing passage of his lecture, Hilbert briefly discussed the connections between analytic and projective geometry. While the theorems and proofs of the former are more general than those of the latter, he said, the methods of the latter are much purer, self-contained, and necessary.[4] By combining synthetic and axiomatic approaches, Hilbert hinted, it should be possible, perhaps, to establish a clear connection between these two branches of the discipline.

In September of that year, Hilbert attended the *Deutsche Mathematiker-Vereinigung* meeting in Halle, where Hermann Wiener (1857–1939) lectured on the foundations of geometry.[5] The lecture could not fail to attract Hilbert's attention given his current teaching interests. Blumenthal reported in 1935 that Hilbert came out greatly excited by what he had just heard, and made his famous declaration that it must be possible to replace "point, line, and plane" with "table, chair, and beer mug" without thereby changing the validity of the theorems of geometry (Blumenthal 1935, 402–403). Seen from the point of view of later developments and what came to be considered the innovative character of *Grundlagen der Geometrie,* this may have been indeed a reason for Hilbert's enthusiasm following the lecture. If we also recall the main points of interest in his 1891 lectures, however, we can assume that Wiener's claim about the possibility of proving central theorems of projective geometry without continuity considerations exerted no lesser impact, and perhaps even a greater one, on Hilbert at the time. Moreover, the idea of changing names of the central concepts while leaving the deductive structure intact was an idea that Hilbert already knew, if not from other, earlier mathematical sources, then at least from his attentive reading of the relevant passages in Dedekind's *Was sind und was sollen die Zahlen?*,[6] where he may not have failed to see the introductory remarks on the role of continuity in geometry. If Hilbert's famous declaration was actually pronounced for the first time after this lecture, as Blumenthal reported, one can then perhaps conclude that Wiener's ideas were more than just a revelation for Hilbert, but acted as a catalyst binding together several threads that may have already been present in his mind for a while.

Roughly at the time when Hilbert's research efforts started to focus on the theory of algebraic number fields, from 1893 on, his interest regarding the foundations of geometry also became more intensive, at least at the level of teaching. In preparing a course on non-Euclidean geometry to be taught that year, Hilbert was already adopting a more axiomatic perspective. The original manuscript of the course clearly reveals that Hilbert had decided to follow more closely the model put forward by Pasch. As for the latter, using the axiomatic approach was a direct expression of a nat-

4 Cf. (Toepell 1986, 37).

5 He may have also attended Wiener's second lecture in 1893. Cf. (Rowe 1999, 556).

6 As we know from a letter to Paul du Bois-Reymond of March-April, 1888. Cf. (Dugac 1976, 203).

uralistic approach to geometry, rather than a formalistic one: the axioms of geometry—Hilbert wrote—express observations of facts of experience, which are so simple that they need no additional confirmation by physicists in the laboratory.[7] From his correspondence with Felix Klein (1849–1925),[8] however, we learn that Hilbert soon realized certain shortcomings in Pasch's treatment, and in particular, certain redundancies that affected it. Hilbert explicitly stipulated at this early stage that a successful axiomatic analysis should aim to establish the *minimal* set of presuppositions from which the whole of geometry could be deduced. Such a task had not been fully accomplished by Pasch himself, Hilbert pointed out, since his Archimedean axiom, could be derived from others in his system.

Hilbert's correspondence also reveals that he kept thinking about the correct way to implement an axiomatic analysis of geometry. In a further letter to Klein, on 15 November while criticizing Lie's approach to the foundations of geometry, he formulated additional tasks to be accomplished by such an analysis. He thus wrote:

> It seems to me that Lie always introduces into the issue a preconceived one-sidedly analytic viewpoint and forgets completely the principal task of non-Euclidean geometry, namely, that of constructing the various possible geometries by the successive introduction of elementary axioms, up until the final construction of the only remaining one, Euclidean geometry.[9]

The course on non-Euclidean geometry was not taught as planned in 1893, since only one student registered for it.[10] It did take place the following year, announced as "Foundations of Geometry." Hilbert had meanwhile considerably broadened his reading in the field, as indicated by the list of almost forty references mentioned in the notes. This list included most of the recent, relevant foundational works. A clear preference for works that followed an empiricist approach is evident, but also articles presenting the ideas of Grassmann were included.[11] It is not absolutely clear to what extent Hilbert read Italian, but none of the current Italian works were included in his list, except for a translated text of Peano (being the only one by a non-German author).[12] It seems quite certain, at any rate, that Hilbert was unaware of the recent works of Fano, Veronese, and others, works that could have been of great interest for him in the direction he was now following.

---

7 "Das Axiom entspricht einer Beobachtung, wie sich leicht durch Kugeln, Lineal und Pappdeckel zeigen lässt. Doch sind diese Erfahrungsthatsachen so einfach, von Jedem so oft beobachtet und daher so bekannt, dass der Physiker sie nicht extra im Laboratorium bestätigen darf." (Hilbert 1893–1894, 10)

8 Hilbert to Klein, 23 May 1893. Quoted in (Frei 1985, 89–90).

9 Hilbert to Klein, 15 November 1893. Quoted in (Frei 1985, 101). On 11 November, he wrote an almost identical letter to Lindemann. Cf. (Toepell 1986, 47).

10 Cf. (Toepell 1986, 51).

11 The full bibliographical list appears in (Toepell 1986, 53–55).

12 At the 1893 annual meeting of the *Deutsche Mathematiker- Vereinigung* in Lübeck (16–20 September), Frege discussed Peano's conceptual language. If not earlier than that, Hilbert certainly heard about Peano's ideas at this opportunity, when he and Minkowski also presented the plans for their expected reports on the theory of numbers. Cf. *Jahresbericht der Deutschen Mathematiker-Vereinigung,* Vol. 4 (1894–1895), p. 8.

Hilbert became acquainted with Hertz's book on the foundations of mechanics, though it was not mentioned in the list. This book seems to have provided a final, significant catalyst for the wholehearted adoption of the axiomatic perspective for geometry. Simultaneously the book established, in Hilbert's view, a direct connection between the latter and the axiomatization of physics in general. Moreover, Hilbert adopted Hertz's more specific, methodological ideas about what is actually involved in axiomatizing a theory. The very fact that Hilbert came to hear about Hertz is not surprising; he would probably have read Hertz's book sooner or later. But that he read it so early was undoubtedly due to Minkowski. During his Bonn years, Minkowski felt closer to Hertz and to his work than to anyone else, and according to Hilbert, his friend had explicitly declared that, had it not been for Hertz's untimely death, he would have dedicated himself exclusively to physics.[13]

Just as with many other aspects of Hilbert's early work, there is every reason to believe that Minkowski's enthusiasm for Hertz was transmitted to his friend. When revising the lecture notes for his course, Hilbert added the following comment:

> Nevertheless the origin [of geometrical knowledge] is in experience. The axioms are, as Hertz would say, pictures or symbols in our mind, such that consequents of the images are again images of the consequences, i.e., what we can logically deduce from the images is itself valid in nature.[14]

Hilbert defined the task to be pursued as part of the axiomatic analysis, including the need to establish the independence of the axioms of geometry. In doing so, however, he stressed once again the objective and factual character of this science. Hilbert wrote:

> The problem can be formulated as follows: What are the necessary, sufficient, and mutually independent conditions that must be postulated for a system of things, in order that any of their properties correspond to a geometrical fact and, conversely, in order that a complete description and arrangement of all the geometrical facts be possible by means of this system of things.[15]

But already at this point it is absolutely clear that, for Hilbert, such questions were not just abstract tasks. Rather, he was directly focused on important, open problems of the discipline, and in particular, on the role of the axiom of continuity in the questions of coordinatization and metrization in projective geometry, as well as in the proof of the fundamental theorems. In a passage that was eventually crossed out, Hilbert expressed his doubts about the prospects of actually proving Wiener's assertion that continuity considerations could be circumvented in projective geometry (Toepell

13 See (Hilbert 1932–1935, 3: 355). Unfortunately, there seems to be no independent confirmation of Minkowski's own statement to this effect.

14 "Dennoch der Ursprung aus der Erfahrung. Die Axiome sind, wie Herz [sic] sagen würde, Bilde[r] oder Symbole in unserem Geiste, so dass Folgen der Bilder wieder Bilder der Folgen sind d.h. was wir aus den Bildern logisch ableiten, stimmt wieder in der Natur." It is worth noting that Hilbert's quotation of Hertz, drawn from memory, was somewhat inaccurate. I am indebted to Ulrich Majer for calling my attention to this passage. (Hilbert 1893–1894, 10)

15 Quoted from the original in (Toepell 1986, 58–59).

1986, 78). Eventually, however, a main achievement of *Grundlagen der Geometrie* would be a detailed realization of this possibility and its consequences, but Hilbert probably decided to follow this direction only after hearing about the result of Friedrich Schur (1856–1932) in 1898. I return to this matter in the next section.

Concerning the validity of the parallel axiom, Hilbert adopted in 1893–1894 a thoroughly empirical approach that reminds us very much of Riemann's *Habilitationsschrift*. Hilbert referred also directly to Gauss's experimental measurement of the sum of angles of the triangle described by three Hannoverian mountain peaks.[16] Although Gauss's measurements were convincing enough for Hilbert to indicate the correctness of Euclidean geometry as a true description of physical space, he still saw an open possibility that future measurements would show it to be otherwise. Hilbert also indicated that existing astronomical observations are not decisive in this respect, and therefore the parallel axiom must be taken at least as a limiting case. In his later lectures on physics, Hilbert would return to this example very often to illustrate the use of axiomatics in physics. In the case of geometry, this particular axiom alone might be susceptible to change following possible new experimental discoveries. Thus, what makes geometry especially amenable to a full axiomatic analysis is the very advanced stage of development it has attained, rather than any other specific, essential trait concerning its nature. In all other respects, geometry is like any other natural science. Hilbert thus stated:

> Among the appearances or facts of experience manifest to us in the observation of nature, there is a peculiar type, namely, those facts concerning the outer shape of things. Geometry deals with these facts. ... Geometry is a science whose essentials are developed to such a degree, that all its facts can already be logically deduced from earlier ones. Much different is the case with the theory of electricity or with optics, in which still many new facts are being discovered. Nevertheless, with regards to its origins, geometry is a natural science.[17]

It is the very process of axiomatization that transforms the natural science of geometry, with its factual, empirical content, into a pure mathematical science. There is no apparent reason why a similar process might not be applied to any other natural science. And in fact, from very early on Hilbert made it clear that this should be done. In the manuscript of his lectures we read that "all other sciences—above all mechanics, but subsequently also optics, the theory of electricity, etc.—should be treated according to the model set forth in geometry."[18]

---

16 The view that Gauss considered his measurement as related to the question of the parallel axiom has been questioned in (Breitenberger 1984) and (Miller 1972). They have argued that this measurement came strictly as a part of Gauss's geodetic investigations. For replies to this argument, see (Scholz 1993, 642–644), and a more recent and comprehensive discussion in (Scholz 2004). Hilbert, at any rate, certainly believed that this had been Gauss's actual intention, and he repeated this opinion on many occasions.

17 Quoted in (Toepell 1986, 58).

18 Quoted in (Toepell 1986, 94).

By 1894, then, Hilbert's interest in foundational issues of geometry had increased considerably, and he had embarked more clearly in an axiomatic direction. His acquaintance with Hertz's ideas helped him conceive the axiomatic treatment of geometry as part of a larger enterprise, relevant also for other physical theories. It also offered methodological guidelines for actually implementing this analysis. However, many of the most important foundational problems remained unsettled for him, and in this sense, even the axiomatic approach did not seem to him to be of great help. At this stage he saw in the axiomatic method no more than an exercise in adding or deleting basic propositions and guessing the consequences that would follow, but certainly not a tool for achieving real new results.[19]

1.2 Geometry in Göttingen

Hilbert moved to Göttingen in 1895 and thereafter he dedicated himself almost exclusively to number theory both in his research and in his teaching. It is worth pointing out, that some of the ideas he developed in this discipline would prove to be essential some years later for his treatment of geometry as presented in *Grundlagen der Geometrie*. In particular, Hilbert's work on the representation of algebraic forms as sums of squares, which had a deep influence on the subsequent development of the theory of real fields,[20] also became essential for Hilbert's own ideas on geometrical constructivity as manifest in *Grundlagen der Geometrie*.

In the summer semester of 1899, Hilbert once again taught a course on the elements of Euclidean geometry. The elaboration of these lectures would soon turn into the famous *Grundlagen der Geometrie*. The very announcement of the course came as a surprise to many in Göttingen, since it signified, on the face of it, a sharp departure from the two fields in which he had excelled since completing his dissertation in 1885: the theory of algebraic invariants and the theory of algebraic number fields. As Blumenthal recalled many years later:

> [The announcement] aroused great excitement among the students, since even the veteran participants of the 'number theoretical walks' (*Zahlkörpersspaziergängen*) had never noticed that Hilbert occupied himself with geometrical questions. He spoke to us only about fields of numbers. (Blumenthal 1935, 402)

Also Hermann Weyl (1855–1955) repeated this view in his 1944 obituary:

> [T]here could not have been a more complete break than the one dividing Hilbert's last paper on the theory of number fields from his classical book *Grundlagen der Geometrie*. (Weyl 1944, 635)

As already suggested, however, the break may have been less sharp than it appeared in retrospect to Hilbert's two distinguished students. Not only because of the strong connections of certain, central results of *Grundlagen der Geometrie* to Hil-

19 As expressed in a letter to Hurwitz, 6 June 1894. See (Toepell 1986, 100).

20 Cf. (Sinaceur 1984, 271–274; 1991, 199–254).

bert's number-theoretical works, or because of Hilbert's earlier geometry courses in Königsberg, but also because Hilbert became actively and intensely involved in current discussions on the foundations of projective geometry starting in early 1898. In fact, at that time Hilbert had attended a lecture in Göttingen given by Schoenflies who discussed a result recently communicated by Schur to Klein, according to which Pappus's theorem could be proven starting from the axioms of congruence alone, and therefore without relying on continuity considerations.[21] Encouraged by this result, and returning to questions that had been raised when he taught the topic several years earlier, Hilbert began to elaborate on this idea in various possible alternative directions. At some point, he even thought, erroneously as it turned out, to have proved that it would suffice to assume Desargues's theorem in order to prove Pappus's theorem.[22]

Schur's result provided the definitive motivation that led Hilbert to embark on an effort to elucidate in detail the fine structure of the logical interdependence of the various fundamental theorems of projective and Euclidean geometry and, more generally, of the structure of the various kinds of geometries that can be produced under various sets of assumptions. The axiomatic method, whose tasks and basic tools Hilbert had been steadily pondering, would now emerge as a powerful and effective instrument for properly addressing these important issues.

The course of 1899 contains much of what will appear in *Grundlagen der Geometrie*. It is worth pointing out here that in the opening lecture Hilbert stated once again the main achievement he expected to obtain from an axiomatic analysis of the foundations of geometry: a complete description, by means of independent statements, of the basic facts from which all known theorems of geometry can be derived. This time he also mentioned the precise source from which this formulation had been taken: the introduction to Hertz's *Principles of Mechanics*.[23] In Hilbert's view, this kind of task was not limited to geometry, and of course also applied, above all, to mechanics. Hilbert had taught seminars on mechanics jointly with Klein in 1897–1898. In the winter semester 1898–1899, he also taught his first full course on a physical topic in Göttingen: mechanics.[24] In the introduction to this course, he explicitly stressed the essential affinity between geometry and the natural sciences, and also explained the role that axiomatization should play in the mathematization of the latter. He compared the two domains in the following terms:

> Geometry also [like mechanics] emerges from the observation of nature, from experience. To this extent, it is an *experimental science*. ... But its experimental foundations are

21 Later published as (Schur 1898).

22 Cf. (Toepell 1986, 114–122). Hessenberg (1905) proves that, in fact, it is Pappus's theorem that implies Desargues's, and not the other way round.

23 Cf. (Toepell 1986, 204).

24 According to the *Nachlass* David Hilbert (Niedersächsische Staats- und Universitätsbibliothek Göttingen, Abteilung Handschriften und Seltene Drucke), (Cod. Ms. D. Hilbert, 520), which contains a list of Hilbert's lectures between 1886 and 1932 (handwritten by Hilbert himself up until 1917–1918), among the earliest courses taught by Hilbert in Königsberg was one in hydrodynamics (summer semester, 1887).

> so irrefutably and so *generally acknowledged*, they have been confirmed to such a degree, that no further proof of them is deemed necessary. Moreover, all that is needed is to derive these foundations from a minimal set of *independent axioms* and thus to construct the whole edifice of geometry by *purely logical means*. In this way [i.e., by means of the axiomatic treatment] geometry is turned into a pure mathematical science. In mechanics it is also the case that all physicists recognize its most *basic facts*. But the *arrangement* of the basic concepts is still subject to a change in perception... and therefore mechanics cannot yet be described today as a *pure mathematical* discipline, at least to the same extent that geometry is. We must strive that it becomes one. We must ever stretch the limits of pure mathematics wider, on behalf not only of our mathematical interest, but rather of the interest of science in general.[25]

This is perhaps the first explicit presentation of Hilbert's program for axiomatizing natural science in general. The more definitive status of the results of geometry, as compared to the relatively uncertain one of our knowledge of mechanics, clearly recalls similar claims made by Hertz. The difference between geometry and other physical sciences—mechanics in this case—was not for Hilbert one of essence, but rather one of historical stage of development. He saw no reason in principle why an axiomatic analysis of the kind he was then developing for geometry could not eventually be applied to mechanics with similar, useful consequences. Eventually, that is to say, when mechanics would attain a degree of development equal to geometry, in terms of the quantity and certainty of known results, and in terms of an appreciation of what really are the "basic facts" on which the theory is based.

2. GRUNDLAGEN DER GEOMETRIE

When Hilbert published his 1899 *Festschrift* (Hilbert 1899) he was actually contributing a further link to a long chain of developments in the foundations of geometry that spanned several decades over the nineteenth century. His works on invariant theory and number theory can be described in similar terms, each within its own field of relevance. In these two fields, as in the foundations of geometry, Hilbert's contribution can be characterized as the "critical" phase in the development of the discipline: a phase in which the basic assumptions and their specific roles are meticulously inspected in order to revamp the whole structure of the theory on a logically sound

25 "Auch die Geometrie ist aus der Betrachtung der Natur, aus der Erfahrung hervorgegangen und insofern eine *Experimentalwissenschaft*. ... Aber diese experimentellen Grundlagen sind so unumstösslich und so *allgemein anerkannt,* haben sich so überall bewährt, dass es einer weiteren experimentellen Prüfung nicht mehr bedarf und vielmehr alles darauf ankommt diese Grundlagen auf ein geringstes Mass *unabhängiger Axiome* zurückzuführen und hierauf *rein logisch* den ganzen Bau der Geometrie aufzuführen. Also Geometrie ist dadurch eine rein *mathematische* Wiss. geworden. Auch in der Mechanik werden die *Grundthatsachen* von allen Physikern zwar anerkannt. Aber die *Anordnung* der Grundbegriffe ist dennoch dem Wechsel der Auffassungen unterworfen... so dass die Mechanik auch heute noch nicht, jedenfalls nicht in dem Masse wie die Geometrie als eine *rein mathematische* Disciplin zu bezeichnen ist. Wir müssen streben, dass sie es wird. Wir müssen die Grenzen echter Math. immer weiter ziehen nicht nur in unserem math. Interesse sondern im Interesse der Wissenschaft überhaupt." (Hilbert 1898–1899, 1–3)

basis and within a logically transparent deductive structure. This time, however, Hilbert had consolidated the critical point of view into an elaborate approach with clearly formulated aims, and affording the proper tools to achieve those aims, at least partly. This was the axiomatic approach that characterizes *Grundlagen der Geometrie* and much of his work thereafter, particularly his research on the foundations of physical theories. However, *Grundlagen der Geometrie* was innovative not only at the methodological level. It was, in fact, a seminal contribution to the discipline, based on a purely synthetic, completely new approach to arithmetizing the various kinds of geometries. And again, as in his two previous fields of research, Hilbert's in-depth acquaintance with the arithmetic of fields of algebraic numbers played a fundamental role in his achievement.

It is important to bear in mind that, in spite of the rigor required for the axiomatic analysis underlying *Grundlagen der Geometrie*, many additions, corrections and improvements—by Hilbert himself, by some of his collaborators and by other mathematicians as well— were still needed over the following years before the goals of this demanding project could be fully attained. Still most of these changes, however important, concerned only the details. The basic structure, the groups of axioms, the theorems considered, and above all, the innovative methodological approach implied by the treatment, all these remained unchanged through the many editions of *Grundlagen der Geometrie*.

The motto of the book was a quotation taken from Kant's *Critique of Pure Reason*: "All human knowledge thus begins with intuitions, proceeds thence to concepts and ends with ideas." If he had to make a choice, Kant appears an almost obvious one for Hilbert in this context. It is hard to state precisely, however, to what extent he had had the patience to become really acquainted with the details of Kant's exacting works. Beyond the well-deserved tribute to his most distinguished fellow Königsberger, this quotation does not seem to offer a reference point for better understanding Hilbert's ideas on geometry.

Hilbert described the aim of his *Festschrift* as an attempt to lay down a "simple" and "complete" system of "mutually independent" axioms, from which all known theorems of geometry might be deduced. His axioms are formulated for three systems of undefined objects named "points," "lines," and "planes," and they establish mutual relations that these objects must satisfy. The axioms are divided into five groups: axioms of incidence, of order, of congruence, of parallels, and of continuity. From a purely logical point of view, the groups have no real significance in themselves. However, from the geometrical point of view they are highly significant, for they reflect Hilbert's actual conception of the axioms as an expression of spatial intuition: each group expresses a particular way that these intuitions manifest themselves in our understanding.

2.1 Independence, Simplicity, Completeness, Consistency

Hilbert's first requirement, that the axioms be independent, is the direct manifestation of the foundational concerns that directed his research. When analyzing independence, his interest focused mainly on the axioms of congruence, continuity and of parallels, since this independence would specifically explain how the various basic theorems of Euclidean and projective geometry are logically interrelated. But as we have seen, this requirement had already appeared—albeit more vaguely formulated—in Hilbert's early lectures on geometry, as a direct echo of Hertz's demand for appropriateness. In *Grundlagen der Geometrie*, the requirement of independence not only appeared more clearly formulated, but Hilbert also provided the tools to prove systematically the mutual independence among the individual axioms within the groups and among the various groups of axioms in the system. He did so by introducing the method that has since become standard: he constructed models of geometries that fail to satisfy a given axiom of the system but satisfy all the others. However, this was not for Hilbert an exercise in analyzing abstract relations among systems of axioms and their possible models. The motivation for enquiring about the mutual independence of the axioms remained, essentially, a geometrical one. For this reason, Hilbert's original system of axioms was not the most economical one from the logical point of view. Indeed, several mathematicians noticed quite soon that Hilbert's system of axioms, seen as a single collection rather than as a collection of five groups, contained a certain degree of redundancy.[26] Hilbert's own aim was to establish the interrelations among the groups of axioms, embodying the various manifestations of special intuition, rather than among individual axioms belonging to different groups.

The second requirement, simplicity, complements that of independence. It means, roughly, that an axiom should contain "no more than a single idea." This is a requirement that Hertz also had explicitly formulated, and Hilbert seemed to be repeating it in the introduction to his own book. Nevertheless, it was neither formally defined nor otherwise realized in any clearly identifiable way within *Grundlagen der Geometrie*. The ideal of formulating "simple" axioms as part of this system was present implicitly as an aesthetic desideratum that was not transformed into a mathematically controllable feature.[27]

The "completeness" that Hilbert demanded for his system of axioms should not be confused with the later, model-theoretical notion that bears the same name, a

---

26 Cf., for instance, (Schur 1901). For a more detailed analysis of this issue, see (Schmidt 1933, 406–408). It is worth pointing out that in the first edition of *Grundlagen der Geometrie* Hilbert stated that he intended to provide an independent system of axioms for geometry. In the second edition, however, this statement no longer appeared, following a correction by E. H. Moore (1902) who showed that one of the axioms might be derived from the others. See also (Corry 2003, §3.5; Torretti 1978, 239 ff.).

27 In a series of articles published in the USA over the first decade of the twentieth century under the influence of *Grundlagen der Geometrie*, see (Corry 2003, §3.5), a workable criterion for simplicity of axioms was systematically sought after. For instance, Edward Huntington (1904, p. 290) included simplicity among his requirements for axiomatic systems, yet he warned that "the idea of a simple statement is a very elusive one which has not been satisfactorily defined, much less attained."

notion that is totally foreign to Hilbert's axiomatic approach at this early stage. Rather it is an idea that runs parallel to Hertz's demand for "correctness." Thus, Hilbert demanded from any adequate axiomatization that it should allow for a derivation of all the known theorems of the discipline in question. The axioms formulated in *Grundlagen der Geometrie*, Hilbert claimed, would indeed yield all the known results of Euclidean geometry or of the so-called absolute geometry, namely that valid independently of the parallel postulate, if the corresponding group of axioms is ignored. Thus, reconstructing the very ideas that had given rise to his own conception, Hilbert discussed in great detail the role of each of the groups of axioms in the proofs of two crucial results: the theorem of Desargues and the theorem of Pappus. Unlike independence, however, the completeness of the system of axioms is not a property that Hilbert knew how to verify formally, except to the extent that, starting from the given axioms, he could prove all the theorems he was interested in.

The question of consistency of the various kinds of geometries was an additional concern of Hilbert's analysis, though, perhaps somewhat surprisingly, one that was not even explicitly mentioned in the introduction to *Grundlagen der Geometrie*. He addressed this issue in the *Festschrift* right after introducing all the groups of axioms and after discussing their immediate consequences. Seen from the point of view of Hilbert's later metamathematical research and the developments that followed it, the question of consistency might appear as the most important one undertaken back in 1899; but in the historical context of the evolution of his ideas it certainly was not. In fact, consistency of the axioms is discussed in barely two pages, and it is not immediately obvious why Hilbert addressed it at all. It doesn't seem likely that in 1899 Hilbert would have envisaged the possibility that the body of theorems traditionally associated with Euclidean geometry might contain contradictions. After all, he conceived Euclidean geometry as an empirically motivated discipline, turned into a purely mathematical science after a long, historical process of evolution and depuration. Moreover, and more importantly, Hilbert had presented a model of Euclidean geometry over certain, special types of algebraic number fields. If with the real numbers the issue of continuity might be thought to raise difficulties that called for particular care, in this case Hilbert would have no real reason to call into question the possible consistency of these fields of numbers. Thus, to the extent that Hilbert referred here to the problem of consistency, he seems in fact to be echoing here Hertz's demand for the permissibility of images. As seen above, a main motivation leading Hertz to introduce this requirement was the concern about possible contradictions brought about over time by the gradual addition of ever new hypotheses to a given theory. Although this was not likely to be the case for the well-established discipline of geometry, it might still have happened that the particular way in which the axioms had been formulated in order to account for the theorems of this science would have led to statements that contradict each other. The recent development of non-Euclidean geometries made this possibility only more patent. Thus, Hilbert believed that, although contradictions might in principle possibly occur within his own system, he could also easily show that this was actually not the case.

The relatively minor importance conceded by Hilbert in 1899 to the problem of the consistency of his system of axioms for Euclidean geometry is manifest not only in the fact that he devoted just two pages to it. Of course, Hilbert could not have in mind a direct proof of consistency here, but rather an indirect one, namely, a proof that any contradiction existing in Euclidean geometry must manifest itself in the arithmetic system of real numbers. This would still leave open the question of the consistency of the latter, a problem difficult enough in itself. However, even an indirect proof of this kind does not appear in explicit form in *Grundlagen der Geometrie*. Hilbert only suggested that it would suffice to show that the specific kind of synthetic geometry derivable from his axioms could be translated into the standard Cartesian geometry, taking the axes as representing the whole field of real numbers.[28] More generally stated, in this first edition of *Grundlagen der Geometrie*, Hilbert preferred to bypass a systematic treatment of the questions related to the structure of the system of real numbers. Rather, he contented himself with constructing a model of his system based on a countable, proper sub-field—of whose consistency he may have been confident—and not the whole field of real numbers (Hilbert 1899, 21). It was only in the second edition of *Grundlagen der Geometrie*, published in 1903, that he added an additional axiom, the so-called "axiom of completeness" (*Vollständigkeitsaxiom*), meant to ensure that, although infinitely many incomplete models satisfy all the other axioms, there is only one complete model that satisfies this last axiom as well, namely, the usual Cartesian geometry, obtained when the whole field of real numbers is used in the model (Hilbert 1903a, 22–24). As Hilbert took pains to stress, this axiom cannot be derived from the Archimedean axiom, which was the only one included in the continuity group in the first edition.[29] It is important to notice, however, that the property referred to by this axiom bears no relation whatsoever to Hilbert's general requirement of "completeness" for any system of axioms. Thus his choice of the term "*Vollständigkeit*" in this context seems somewhat unfortunate.

3. THE 1900 LIST OF PROBLEMS

Soon after the publication of *Grundlagen der Geometrie*, Hilbert had a unique opportunity to present his views on mathematics in general and on axiomatics in particular, when he was invited to address the Second International Congress of Mathematicians

28 And the same is true for Hilbert's treatment of "completeness" (in his current terminology) at that time.

29 The axiom is formulated in (Hilbert 1903a, 16). Toepell (1986, 254–256) briefly describes the relationship between Hilbert's Vollständigkeitsaxiom and related works of other mathematicians. The axiom underwent several changes throughout the various later editions of the Grundlagen, but it remained central to this part of the argument. Cf. (Peckhaus 1990, 29–35). The role of this particular axiom within Hilbert's axiomatics and its importance for later developments in mathematical logic is discussed in (Moore 1987, 109–122). In 1904 Oswald Veblen introduced the term "categorical" (Veblen 1904, 346) to denote a system to which no irredundant axioms may be added. He believed that Hilbert had checked this property in his own system of axioms. See (Scanlan 1991, 994).

held in Paris in August of 1900. The invitation was a definite sign of the reputation that Hilbert had acquired by then within the international mathematics community. Following a suggestion of Minkowski, Hilbert decided to use the opportunity to provide a glimpse into what, in his view, the new century would bring for mathematics. Thus he posed a list of problems that he considered significant challenges that could lead to fruitful research and to new and illuminating ideas for mathematicians involved in solving them.

In many ways, Hilbert's talk embodied his overall vision of mathematics and science, and he built the list of problems to a large extent according to his own mathematical horizons.[30] Some of the problems belonged to number theory and the theory of invariants, the domains that his published work had placed him in among the leading world experts. Some others belonged to domains with which he was closely acquainted, even though he had not by then published anything of the same level of importance, such as variational calculus. It further included topics that Hilbert simply considered should be given a significant push within contemporary research, such as Cantorian set theory. The list reflected Hilbert's mathematical horizon also in the sense that a very significant portion of the works he cited in reference to the various problems had been published in either of the two main Göttingen mathematical venues: the *Mathematische Annalen* and the Proceedings of the Göttingen Academy of Sciences. And although Hilbert's mathematical horizons were unusually broad, they were nonetheless clearly delimited and thus, naturally, several important, contemporary fields of research were left out of the list.[31] Likewise, important contemporary Italian works on geometry, and the problems related to them, were not referred to at all in the geometrical topics that Hilbert did consider in his list. Moreover, two major contemporary open problems, Fermat's theorem and Poincaré's three-body problem, though mentioned in the introduction, were not counted among the twenty-three problems.

The talk also reflected three other important aspects of Hilbert's scientific personality. Above all is his incurable scientific optimism, embodied in the celebrated and often quoted statement that every mathematical problem can indeed be solved: "There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignorabimus*." This was meant primarily as a reaction to a well-known pronouncement of the physiologist Emil du Bois-Reymond (1818–1896) on the inherent limitations of science as a system able to provide us with knowledge about the world.[32] Second, is the centrality of challenging problems in mathematics as a main, necessary condition for the healthy development of any branch of the discipline and, more generally, of that living organism that Hilbert took mathematics to be. And third, is the central role accorded to empirical motivations as a fundamental source of nourishment for that organism, in which mathematics and the physical sci-

30 Several versions of the talk appeared in print and they were all longer and more detailed than the actual talk. Cf. (Grattan-Guinness 2000).

31 Cf. (Gray 2000, 78–88).

ences appear tightly interrelated. But stressing the empirical motivations underlying mathematical ideas should by no means be taken as opposed to rigor. On the contrary, contrasting an "opinion occasionally advocated by eminent men," Hilbert insisted that the contemporary quest for rigor in analysis and arithmetic should in fact be *extended to both geometry and the physical sciences*. He was alluding here, most probably, to Kronecker and Weierstrass, and the Berlin purist tendencies that kept geometry and applications out of their scope of interest. Rigorous methods are often simpler and easier to understand, Hilbert said, and therefore, a more rigorous treatment would only perfect our understanding of these topics, and at the same time would provide mathematics with ever new and fruitful ideas. Explaining why rigor should not be sought only within analysis, Hilbert actually implied that this rigor should actually be pursued in axiomatic terms. He thus wrote:

> Such a one-sided interpretation of the requirement of rigor would soon lead to the ignoring of all concepts arising form geometry, mechanics and physics, to a stoppage of the flow of new material from the outside world, and finally, indeed, as a last consequence, to the rejection of the ideas of the continuum and of irrational numbers. But what an important nerve, vital to mathematical science, would be cut by rooting out geometry and mathematical physics! On the contrary I think that wherever mathematical ideas come up, whether from the side of the theory of knowledge or in geometry, or from the theories of natural or physical science, the problem arises for mathematics to investigate the principles underlying these ideas and to establish them upon a simple and complete system of axioms, so that the exactness of the new ideas and their applicability to deduction shall be in no respect inferior to those of the old arithmetical concepts.[33]

Using rhetoric reminiscent of Paul Volkmann's 1900 book, Hilbert described the development of mathematical ideas as an ongoing, dialectical interplay between the two poles of thought and experience, an interplay that brings to light a "pre-established harmony" between nature and mathematics.[34] The "edifice metaphor" was invoked to help stress the importance of investigating the foundations of mathematics not as an isolated concern, but rather as an organic part of the manifold growth of the discipline in several directions. Hilbert thus said:

> Indeed, the study of the foundations of a science is always particularly attractive, and the testing of these foundations will always be among the foremost problems of the investi-

32 See (Du Bois-Reymond 1872). Hilbert would repeat this claim several times later in his career, notably in (Hilbert 1930). Although the basic idea behind the pronouncement was the same on all occasions, and it always reflected his optimistic approach to the capabilities of mathematics, it would nevertheless be important to consider the specific, historical framework in which the pronouncement came and the specific meaning that the situation conveys in one and the same sentence. If in 1900 it came, partly at least, as a reaction to Du Bois-Reymond's sweeping claim about the limitation of science, in 1930 it came after the intense debate against constructivist views about the foundations of arithmetic.

33 The classical locus for the English version of the talk is (Hilbert 1902a). Here I have preferred to quote, where different, from the updated translation appearing in (Gray 2000, 240–282). This passage appears there on p. 245.

34 The issue of the "pre-established harmony" between mathematics and nature was a very central one among Göttingen scientists. This point has been discussed in (Pyenson 1982).

> gator ... [But] a thorough understanding of its special theories is necessary for the successful treatment of the foundations of the science. Only that architect is in the position to lay a sure foundation for a structure who knows its purpose thoroughly and in detail.[35]

Speaking more specifically about the importance of problems for the healthy growth of mathematics, Hilbert characterized an interesting problem as one that is "difficult in order to entice us, yet not completely inaccessible, lest it mock our efforts." But perhaps more important was the criterion he formulated for the solution of one such problem: it must be possible "to establish the correctness of the solution by a finite number of steps based upon a finite number of hypotheses which are implied in the statement of the problem and which must always be exactly formulated."

3.1 Foundational Problems

This is not the place to discuss in detail the list of problems and their historical background and development.[36] Our main concern here is with the sixth problem— Hilbert's call for the axiomatization of physical sciences—and those other problems on the list more directly connected with it. The sixth problem is indeed the last of a well-defined group within the list, to which other "foundational" problems also belong. Beyond this group, the list can be said roughly to contain three other main areas of interest: number theory, algebraic-geometrical problems, and analysis (mainly variational calculus) and its applications in physics.

The first two foundational problems, appearing at the head of Hilbert's list, are Cantor's continuum hypothesis and the compatibility of the axioms of arithmetic. In formulating the second problem on his list, Hilbert stated more explicitly than ever before, that among the tasks related to investigating an axiomatic system, proving its consistency would be the most important one. Eventually this turned into a main motto of his later program for the foundations of arithmetic beginning in the 1920s, but many years and important developments still separated this early declaration, diluted among a long list of other important mathematical tasks for the new century, from an understanding of the actual implications of such an attempt and from an actual implementation of a program to pursue it. In the years to come, as we will see below, Hilbert did many things with axiomatic systems other than attempting a proof of consistency for arithmetic.

Hilbert stated that proving the consistency of geometry could be reduced to proving that of arithmetic, and that the axioms of the latter were those presented by him in "Über den Zahlbegriff" several months prior to this talk. Yet, Hilbert was still confident that this would be a rather straightforward task, easily achievable "by means of a careful study and suitable modification of the known methods of reasoning in the theory of irrational numbers" (Hilbert 1902a, 448).Hilbert did not specify the exact

35 Quoted from (Gray 2000, 258).

36 Cf. (Rowe 1996), and a more detailed, recent, discussion in (Gray 2000).

meaning of this latter statement, but its wording would seem to indicate that in the system of axioms proposed for arithmetic, the difficulty in dealing with consistency would come from the assumption of continuity. Thus the consistency of Euclidean geometry would depend on proving the consistency of arithmetic as defined by Hilbert through his system of axioms. This would, moreover, provide a proof for the very existence of the continuum of real numbers as well. Clearly Hilbert meant his remarks in this regard to serve as an argument against Kronecker's negative reactions to unrestricted use of infinite collections in mathematics, and therefore he explicitly asserted that a consistent system of axioms could prove the existence of higher Cantorian cardinals and ordinals.[37] He thus established a clear connection between the two first problems on his list through the axiomatic approach. Still, Hilbert was evidently unaware of the difficulties involved in realizing this point of view, and, more generally, he most likely had no precise idea of what an elaborate theory of systems of axioms would involve. On reading the first draft of the Paris talk, several weeks earlier, Minkowski understood at once the challenging implications of Hilbert's view, and he hastened to write to his friend:

> In any case, it is highly original to proclaim as a problem for the future, one that mathematicians would think they had already completely possessed for a long time, such as the axioms for arithmetic. What might the many laymen in the auditorium say? Will their respect for us grow? And you will also have a though fight on your hands with the philosophers.[38]

Minkowski turned out to be right to a large extent, and among the ideas that produced the strongest reactions were those related with the status of axioms as implicit definitions, such as Hilbert introduced in formulating the second problem. He thus wrote:

> When we are engaged in investigating the foundations of a science, we must set up a system of axioms which contains an exact and complete description of the relations subsisting between the elementary ideas of the science. The axioms so set up are at the same time the definitions of those elementary ideas, and no statement within the realm of the science whose foundation we are testing is held to be correct unless it can be derived from those axioms by means of a finite number of logical steps. (Hilbert 1902a,447)[39]

The next three problems in the list are directly related with geometry and, although not explicitly formulated in axiomatic terms, they address the question of finding the correct relationship between specific assumptions and specific, significant geometrical facts. Of particular interest for the present account is the fifth. The question of the foundations of geometry had evolved over the last third of the nineteenth century along two parallel paths. First was the age-old tradition of elementary synthetic

---

37 Hilbert also pointed out that no consistent set of axioms could be similarly set up for all cardinals and all alephs. Commenting on this, Ferreirós (1999, 301), has remarked: "This is actually the first published mention of the paradoxes of Cantorian set theory — without making any fuss of it." See also (Peckhaus and Kahle 2002).

38 On 17 July 1900, (Rüdenberg and Zassenhaus 1973, 129).

39 And also quoted in (Gray 2000, 250).

geometry, where the question of foundations more naturally arises in axiomatic terms. A second, alternative, path, that came to be associated with the Helmholtz-Lie problem, had derived directly from the work of Riemann and it had a more physically-grounded orientation connected with the question of spaces that admit the free mobility of rigid bodies. Whereas Helmholtz had only assumed continuity as underlying the motion of rigid bodies, in applying his theory of group of transformations to this problem, Lie was also assuming the differentiability of the functions involved. Hilbert's work on the foundations of geometry, especially in the context that led to *Grundlagen der Geometrie*, had so far been connected with the first of these two approaches, while devoting much less attention to the second one. Now in his fifth problem, he asked whether Lie's conditions, rather than assumed, could actually be deduced from the group concept together with other geometrical axioms.

As a mathematical problem, the fifth one led to interesting, subsequent developments. Not long after his talk, on 18 November 1901, Hilbert himself proved that, in the plane, the answer is positive, and he did so with the help of a then innovative, essentially topological, approach (Hilbert 1902b). That the answer is positive in the general case was satisfactorily proved only in 1952.[40] What concerns us here more directly, however, is that the inclusion of this problem in the list underscores the actual scope of Hilbert's views over the question of the foundations of geometry and over the role of axiomatics. Hilbert suggested here the pursuit of an intricate kind of conceptual clarification involving our assumptions about motion, differentiability and symmetry, such as they appear intimately interrelated in the framework of a well-elaborate mathematical theory, namely, that of Lie. This quest is typical of the spirit of Hilbert's axiomatic involvement with physical theories. At this point, it also clearly suggests that his foundational views on geometry were much broader and open-ended than an exclusive focusing on *Grundlagen der Geometrie* — with a possible overemphasizing of certain, formalist aspects — might seem to imply. In particular, the fifth problem emphasizes, once again and from a different perspective, the prominent role that Hilbert assigned to physicalist considerations in his approach to geometry. In the long run, one can also see this aspect of Hilbert's view resurfacing at the time of his involvement with general theory of relativity. In its more immediate context, however, it makes the passage from geometry to the sixth problem appear as a natural one within the list.

Indeed, if the first two problems in the list show how the ideas deployed in *Grundlagen der Geometrie* led in one direction towards foundational questions in arithmetic, then the fifth problem suggests how they also naturally led, in a different direction, to Hilbert's call for the axiomatization of physical science in the sixth problem. The problem was thus formulated as follows:

> The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays

40 This was done, simultaneously, in (Gleason 1952) and (Montgomery and Zippin 1952).

> an important part; in the first rank are the theory of probabilities and mechanics. (Hilbert 1902a, 454)[41]

As examples of what he had in mind Hilbert mentioned several existing and well-known works: the fourth edition of Mach's *Die Mechanik in ihrer Entwicklung*, Hertz's *Principles*, Boltzmann's 1897 *Vorlesungen Über die Principien der Mechanik*, and also Volkmann's 1900 *Einführung in das Studium der theoretischen Physik*. Boltzmann's work offered a good example of what axiomatization would offer, as he had indicated, though only schematically, that limiting processes could be applied, starting from an atomistic model, to obtain the laws of motion of continua. Hilbert thought it convenient to go in the opposite direction also, i.e., to derive the laws of motions of rigid bodies by limiting processes, starting from a system of axioms that describe space as filled with continuous matter in varying conditions. Thus one could investigate the equivalence of different systems of axioms, an investigation that Hilbert considered to be of the highest theoretical importance.

This is one of the few places where Hilbert emphasized Boltzmann's work over Hertz's in this regard, and this may give us the clue to the most immediate trigger that was in the back of Hilbert's mind when he decided to include this problem in the list. Hilbert had met Boltzmann several months earlier in Munich, where he heard his talk on recent developments in physics. Boltzmann had not only discussed ideas connected to the task that Hilbert was now calling for, but he also adopted a rhetoric that Hilbert seems to have found very much to the point. In fact, Boltzmann had suggested that one could follow up the recent history of physics with a look at future developments. Nevertheless, he said, "I will not be so rash as to lift the veil that conceals the future" (Boltzmann 1899, 79). Hilbert, on the contrary, opened the lecture by asking precisely, "who among us would not be glad to lift the veil behind which the future lies hidden" and the whole trust of his talk implied that he, the optimistic Hilbert, was helping the mathematical community to do so.

Together with the well-known works on mechanics referred to above, Hilbert also mentioned a recent work by the Göttingen actuarial mathematician Georg Bohlmann (1869–1928) on the foundations of the calculus of probabilities.[42] The latter was important for physics, Hilbert said, for its application to the method of mean values and to the kinetic theory of gases. Hilbert's inclusion of the theory of probabilities among the main *physical* theories whose axiomatization should be pursued has often puzzled readers of this passage. It is also remarkable that Hilbert did not mention electrodynamics among the physical disciplines to be axiomatized, even though the second half of the Gauss-Weber *Festschrift*, where Hilbert's *Grundlagen der Geometrie* was published, contained a parallel essay by Emil Wiechert (1861–1956) on the foundations of electrodynamics (Wiechert 1899). At any rate, Wiechert's presentation

---

41 Quoted in (Gray 2000, 257).

42 This article reproduced a series of lectures delivered by Bohlmann in a *Ferienkurs* in Göttingen (Bohlmann 1900). In his article Bohlmann referred the readers, for more details, to the chapter he had written for the *Encyklopädie* on insurance mathematics.

was by no means axiomatic, in any sense of the term. On the other hand, the topics addressed by him would start attracting Hilbert's attention over the next years, at least since 1905.

Modelling this research on what had already been done for geometry meant that not only theories considered to be closer to "describing reality" should be investigated, but also other, logically possible ones. The mathematician undertaking the axiomatization of physical theories should obtain a complete survey of all the results derivable from the accepted premises. Moreover, echoing the concern already found in Hertz and later to appear also in Hilbert's letters to Frege, a main task of the axiomatization would be to avoid that recurrent situation in physical research, in which new axioms are added to existing theories without properly checking to what extent the former are compatible with the latter. This proof of compatibility, concluded Hilbert, is important not only in itself, but also because it compels us to search for ever more precise formulations of the axioms.

3.2 A Context for the Sixth Problem

The sixth problem of the list deals with the axiomatization of physics. It was suggested to Hilbert by his own recent research on the foundations of geometry. He thus proposed "to treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part." This sixth problem is not really a problem in the strict sense of the word, but rather a general task for whose complete fulfilment Hilbert set no clear criteria. Thus, Hilbert's detailed account in the opening remarks of his talk as to what a meaningful problem in mathematics is, and his stress on the fact that a solution to a problem should be attained in a finite number of steps, does not apply in any sense to the sixth one. On the other hand, the sixth problem has important connections with three other problems on Hilbert's list: the nineteenth ("Are all the solutions of the Lagrangian equations that arise in the context of certain typical variational problems necessarily analytic?"), the twentieth (dealing with the existence of solutions to partial differential equations with given boundary conditions), closely related to the nineteenth and at the same time to Hilbert's long-standing interest in the Dirichlet principle,[43] and, finally, the twenty-third (an appeal to extend and refine the existing methods of variational calculus). Like the sixth problem, the latter two are general tasks rather than specific mathematical problems with a clearly identifiable, possible solution.[44] All these three problems are also strongly connected to physics, though unlike the sixth, they are also part of mainstream, tradi-

43 On 11 October 1899, Hilbert had lectured in Göttingen on the Dirichlet principle, stressing the importance of its application to the theory of surfaces and also to mathematical physics. Cf. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8 (1900), 22.

44 A similar kind of "general task" problem that Hilbert had perhaps considered adding as the twenty-fourth problem in his list is hinted at in an undated manuscript found in *Nachlass* David Hilbert (Cod. Ms. D. Hilbert, 600). It concerns the definition of criteria for finding simplest proofs in mathematics in general. Cf. a note in (Grattan-Guinness 2001, 167), and a more detailed account in (Thiele 2003).

tional research concerns in mathematics.[45] In fact, their connections to Hilbert's own interests are much more perspicuous and, in this respect, they do not raise the same kind of historical questions that Hilbert's interest in the axiomatization of physics does. Below, I will explain in greater detail how Hilbert conceived the role of variational principles in his program for axiomatizing physics.

Another central issue to be discussed below in some detail is the role the sixth problem played in subsequent developments in mathematics and in physics. At this stage, however, a general point must be stressed about the whole list in this regard. A balanced assessment of the influence of the problems on the development of mathematics throughout the century must take into account not only the intrinsic importance of the problems,[46] but also the privileged institutional role of Göttingen in the mathematical world with the direct and indirect implications of its special status. If Hilbert wished to influence the course of mathematics over the coming century with his list, then his own career was only very partially shaped by it. Part of the topics covered by the list belonged to his previous domains of research, while others belonged to domains where he never became active. On the contrary, domains that he devoted much effort to over the next years, such as the theory of integral equations, were not contemplated in the list. In spite of the enormous influence Hilbert had on his students, the list did not become a necessary point of reference of preferred topics for dissertations. To be sure, some young mathematicians, both in Göttingen and around the world, did address problems on the list and sometimes came up with important mathematical achievements that helped launch their own international careers. But this was far from the only way for talented young mathematicians to reach prominence in or around Göttingen. But, ironically, the sixth problem, although seldom counted among the most influential of the list, will be shown here to count among those that received a greater attention from Hilbert himself and from his collaborators and students over the following years.

For all its differences and similarities with other problems on the list, the important point that emerges from the above account is that the sixth problem was in no sense disconnected from the evolution of Hilbert's early axiomatic conception. Nor was it artificially added in 1900 as an afterthought about the possible extensions of an idea successfully applied in 1899 to the case of geometry. Rather, Hilbert's ideas concerning the axiomatization of physical science arose simultaneously with his increasing enthusiasm for the axiomatic method and they fitted naturally into his overall view of pure mathematics, geometry and physical science—and the relationship among them—by that time. Moreover, as will be seen in the next chapter in some detail, Hilbert's 1905 lectures on axiomatization provide a very clear and comprehensive conception of how the project suggested in the sixth problem should be realized. In fact, it is very likely that this conception was not essentially different from what Hilbert had in mind when formulating his problem in 1900.[47] Interestingly, the devel-

45 For a detailed account of the place of variational principles in Hilbert's work, see (Blum 1994).

46 As treated in (Alexandrov 1979; Browder 1976).

opment of physics from the beginning of the century, and especially *after* 1905, brought many surprises that Hilbert could not have envisaged in 1900 or even when he lectured at Göttingen on the axioms of physics; yet, over the following years Hilbert was indeed able to accommodate these new developments to the larger picture of physics afforded by his program for axiomatization. In fact, some of his later contributions to mathematical physics came by way of realizing the vision embodied in this program, as will be seen in detail in later chapters.

4. FOUNDATIONAL CONCERNS – EMPIRICIST STANDPOINT

Following the publication of *Grundlagen der Geometrie*, Hilbert was occupied for a while with research on the foundations of geometry. Several of his students, such as Max Dehn (1878–1952), Georg Hamel (1877–1954) and Anne Lucy Bosworth (1868–1907), worked in this field as well, including on problems relating to Hilbert's 1900 list. Also many meetings of the *Göttinger Mathematische Gesellschaft* during this time were devoted to discussing related topics. On the other hand, questions relating to the foundations of arithmetic and set theory also received attention in the Hilbert circle. Ernst Zermelo (1871–1953) had already arrived in Göttingen in 1897 in order to complete his *Habilitation*, and his own focus of interest changed soon from mathematical physics to set theory and logic. Around 1899–1900 he had already found an important antinomy in set theory, following an idea of Hilbert's.[48] Later on, in the winter semester of 1900–1901, Zermelo was teaching set theory in Göttingen (Peckhaus 1990, 48–49).

Interest in the foundations of arithmetic became a much more pressing issue in 1903, after Bertrand Russell (1872–1970) published his famous paradox arising from Frege's logical system. Although Hilbert hastened to indicate to Frege that similar arguments had been known in Göttingen for several years,[49] it seems that Russell's publication, coupled with the ensuing reaction by Frege,[50] did have an exceptional impact. Probably this had to do with the high esteem that Hilbert professed towards Frege's command of these topics (which Hilbert may have come to appreciate even more following the sharp criticism recently raised by the latter towards his own ideas). The simplicity of the sets involved in Russell's argument was no doubt a further factor that explains its strong impact on the Göttingen mathematicians. If Hilbert had initially expected that the difficulty in completing the full picture of his approach to the foundations of geometry would lie on dealing with more complex assumptions such as the *Vollständigkeitsaxiom*, now it turned out that the problems perhaps started with the arithmetic itself and even with logic. He soon realized that greater attention

47 Cf. (Hochkirchen 1999), especially chap. 1.

48 See (Peckhaus and Kahle 2002).

49 Hilbert to Frege, 7 November 1903. Quoted in (Gabriel et al. 1980, 51–52).

50 As published in (Frege 1903, 253). See (Ferreirós 1999, 308–311).

should be paid to these topics, and in particular to the possible use of the axiomatic method in establishing the consistency of arithmetic (Peckhaus 1990, 56–57).

Hilbert himself gradually reduced his direct involvement with all questions of this kind, and after 1905 he completely abandoned them for many years to come. Two instances of his involvement with foundational issues during this period deserve some attention here. The first is his address to the Third International Congress of Mathematicians, held in 1904 in Heidelberg. In this talk, later published under the title of "On the Foundations of Logic and Arithmetic," Hilbert presented a program for attacking the problem of consistency as currently conceived. The basic idea was to develop simultaneously the laws of logic and arithmetic, rather than reducing one to the other or to set theory. The starting point was the basic notion of thought-object that would be designated by a sign, which offered the possibility of treating mathematical proofs, in principle, as formulae. This could be seen to constitute an interesting anticipation of what later developed as part of Hilbert's proof theory, but here he only outlined the idea in a very sketchy way. Actually, Hilbert did not go much beyond the mere declaration that this approach would help achieve the desired proof. Hilbert cursorily reviewed several prior approaches to the foundations of arithmetic, only to discard them all. Instead, he declared that the solution for this problem would finally be found in the correct application of the axiomatic method (Hilbert 1905c, 131).

Upon returning to Göttingen from Heidelberg, Hilbert devoted some time to working out the ideas outlined at the International Congress of Mathematicians. The next time he presented them was in an introductory course devoted to "The Logical Principles of Mathematical Thinking," which contains the second instance of Hilbert's involvement with the foundation of arithmetic in this period. This course is extremely important for my account here because it contains the first detailed attempt to implement the program for the axiomatization of physics.[51] I will examine it in some detail below. At this point I just want to briefly describe the other parts of the course, containing some further foundational ideas for logic and arithmetic, and some further thoughts on the axiomatization of geometry.

Hilbert discussed in this course the "logical foundations" of mathematics by introducing a formalized calculus for propositional logic. This was a rather rudimentary calculus, which did not even account for quantifiers. As a strategy for proving consistency of axiomatic systems, it could only be applied to very elementary cases.[52] Prior to defining this calculus Hilbert gave an overview of the basic principles of the axiomatic method, including a more detailed account of its application to arithmetic, geometry and the natural sciences. What needs to be stressed concerning this text is that, in spite of his having devoted increased attention over the previous years to foundational questions in arithmetic, Hilbert's fundamentally empiricist

51 There are two extant sets of notes for this course: (Hilbert 1905a and 1905b). Quotations below are taken from (Hilbert 1905a). As these important manuscripts remain unpublished, I transcribe in the footnotes some relevant passages at length. Texts are underlined or crossed-out as in the original. Later additions by Hilbert appear between < > signs.

52 For a discussion of this part of the course, see (Peckhaus 1990, 61–75).

approach to issues in the foundations of geometry was by no means weakened, but rather the opposite. In fact, in his 1905 course, Hilbert actually discussed the role of an axiomatic analysis of the foundations of arithmetic in similar, empiricist terms.

Once again, Hilbert contrasted the axiomatic method with the genetic approach in mathematics, this time making explicit reference to the contributions of Kronecker and Weierstrass to the theory of functions. Yet Hilbert clearly separated the purely logical aspects of the application of the axiomatic method from the "genetic" origin of the axioms themselves: the latter is firmly grounded on empirical experience. Thus, Hilbert asserted, it is not the case that the system of numbers is given to us through the network of concepts (*Fachwerk von Begriffen*) involved in the eighteen axioms. On the contrary, it is our direct intuition of the concept of natural number and of its successive extensions, well known to us by means of the genetic method, which has guided our construction of the axioms:

> The aim of every science is, first of all, to set up a network of concepts based on axioms to whose very conception we are naturally led by *intuition and experience*. Ideally, all the phenomena of the given domain will indeed appear as part of the network and all the theorems that can be derived from the axioms will find their expression there.[53]

What this means for the axiomatization of geometry, then, is that its starting point must be given by the intuitive facts of that discipline,[54] and that the latter must be in agreement with the network of concepts created by means of the axiomatic system. The concepts involved in the network itself, Hilbert nevertheless stressed, are totally detached from experience and intuition.[55] This procedure is rather obvious in the case of arithmetic, and to a certain extent the genetic method has attained similar results for this discipline. In the case of geometry, although the need to apply the pro-

---

53 "*Uns war das Zahlensystem schließlich nichts* als ein *Fachwerk von Begriffen*, das durch 18 Axiome definiert war. Bei der Aufstellung dieser leitete uns allerdings die Anschauung, die wir von dem Begriff der Anzahl und seiner genetischen Ausdehnung haben. ... So ist in jeder Wissenschaft die Aufgabe, in den Axiomen zunächst ein Fachwerk von Begriffen zu errichten, bei dessen Aufstellung wir uns natürlich durch die *Anschauung und Erfahrung* leiten lassen; das Ideal ist dann, daß in diesem Fachwerk alle Erscheinungen des betr. Gebietes Platz finden, und daß jeder aus den Axiomen folgende Satz dabei Verwertung findet.
Wollen wir nun für die Geometrie ein Axiomensystem aufstellen, so heißt das, daß wir uns den Anlaß dazu durch die anschaulichen Thatsachen der Geometrie geben lassen, und diesen das aufzurichtende Fachwerk entsprechen lassen; die Begriffe, die wir so erhalten, sind *aber als gänzlich losgelöst von jeder Erfahrung und Anschauung zu betrachten*. Bei der Arithmetik ist diese Forderung verhältnismäßig naheliegend, sie wird in gewissem Umfange auch schon bei der genetischen Methode angestrebt. Bei der Geometrie jedoch wurde die Notwendigkeit dieses Vorgehens viel später erkannt; dann aber wurde eine axiomatische Behandlung eher versucht, als in der Arithmetik, wo noch immer die genetische Betrachtung herrschte. Doch ist die Aufstellung eines vollständigen Axiomensystemes ziemlich schwierig, noch viel schwerer wird sie in der Mechanik, Physik etc. sein, wo das Material an Erscheinungen noch viel größer ist." (Hilbert 1905a, 36–37)

54 "... den Anlaß dazu durch die anschaulischen Thatsachen der Geometrie geben lassen..." (Hilbert 1905a, 37)

55 "... die Begriffe, die wir so erhalten, sind aber als gänzlich losgelöst von jeder Erfahrung und Anschauung zu betrachten." (Hilbert 1905a, 37)

cess truly systematically was recognized much later, the axiomatic presentation has traditionally been the accepted one. And if setting up a full axiomatic system has proven to be a truly difficult task for geometry, then, Hilbert concluded, it will be much more difficult in the case of mechanics or physics, where the range of observed phenomena is even broader.[56]

Hilbert's axioms for geometry in 1905 were based on the system of *Grundlagen der Geometrie*, including all the corrections and additions introduced to it since 1900. Here too he started by choosing three basic kinds of undefined elements: points, lines and planes. This choice, he said, is somewhat "arbitrary" and it is dictated by consideration of simplicity. But the arbitrariness to which Hilbert referred here has little to do with the arbitrary choice of axioms sometimes associated with twentieth-century formalistic conceptions of mathematics; it is not an absolute arbitrariness constrained only by the requirement of consistency. On the contrary, it is limited by the need to remain close to the "intuitive facts of geometry." Thus, Hilbert said, instead of the three chosen, basic kinds of elements, one could likewise start with [no... not with "chairs, tables, and beer-mugs," but rather with] circles and spheres, and formulate the adequate axioms that are still in agreement with the usual, intuitive geometry.[57]

Hilbert plainly declared that Euclidean geometry—as defined by his systems of axioms—is the one and only geometry that fits our spatial experience,[58] though in his opinion, it would not be the role of mathematics or logic to explain why this is so. But if that is the case, then what is the status of the non-Euclidean or non-Archimedean geometries? Is it proper at all to use the term "geometry" in relation to them? Hilbert thought it unnecessary to break with accepted usage and restrict the meaning of the term to cover only the first type. It has been unproblematic, he argued, to extend the meaning of the term "number" to include also the complex numbers, although the latter certainly do not satisfy all the axioms of arithmetic. Moreover, it would be untenable from the logical point of view to apply the restriction: although it is not highly probable, it may nevertheless be the case that some changes would still be introduced in the future to the system of axioms that describes intuitive geometry. In fact, Hilbert knew very well that this "improbable" situation had repeatedly arisen in relation to the original system he had put forward in 1900 in *Grundlagen der Geometrie*. To conclude, he compared once again the respective situations in geometry and in physics: in the theory of electricity, for instance, new theories are continually formulated that transform many of the basic facts of the discipline, but no one thinks that the name of the discipline needs to be changed accordingly.

---

56 "... das Material an Erscheinungen noch viel größer ist." (Hilbert 1905a, 37)

57 "Daß wir gerade diese zu Elementardingen des begrifflichen Fachwerkes nehmen, ist willkürlich und geschieht nur wegen ihrer augenscheinlichen Einfachheit; im Princip könnte man die ersten Dinge auch Kreise und Kugeln nennen, und die Festsetzungen über sie so treffen, daß sie diesen Dingen der anschaulichen Geometrie entsprechen." (Hilbert 1905a, 39)

58 "Die Frage, wieso man in der Natur nur gerade die durch alle diese Axiome festgelegte Euklidische Geometrie braucht, bezw. warum unsere Erfahrung gerade in dieses Axiomsystem sich einfügt, gehört nicht in unsere mathematisch-logichen Untersuchungen." (Hilbert 1905a, 67)

Hilbert also referred explicitly to the status of those theories that, like non-Euclidean and non-Archimedean geometries, are created arbitrarily through the purely logical procedure of setting down a system of independent and consistent axioms. These theories, he said, can be applied to any objects that satisfy the axioms. For instance, non-Euclidean geometries are useful to describe the paths of light in the atmosphere under the influence of varying densities and diffraction coefficients. If we assume that the speed of light is proportional to the vertical distance from a horizontal plane, then one obtains light-paths that are circles orthogonal to the planes, and light-times equal to the non-Euclidean distance from them.[59] Thus, the most advantageous way to study the relations prevailing in this situation is to apply the conceptual schemes provided by non-Euclidean geometry.[60]

A further point of interest in Hilbert's discussion of the axioms of geometry in 1905 concerns his remarks about what he called the philosophical implications of the use of the axiomatic method. These implications only reinforced Hilbert's empiricist view of geometry. Geometry, Hilbert said, arises from reality through intuition and observation, but it works with idealizations: for instance, it considers very small bodies as points. The axioms in the first three groups of his system are meant to express idealizations of a series of facts that are easily recognizable as independent from one other; the assertion that a straight line is determined by two points, for instance, never gave rise to the question of whether or not it follows from other, basic axioms of geometry. But establishing the status of the assertion that the sum of the angles in a triangle equals two right angles requires a more elaborate axiomatic analysis. This analysis shows that such an assertion is a separate piece of knowledge, which—we now know for certain—cannot be deduced from earlier facts (or from their idealizations, as embodied in the three first groups of axioms). This knowledge can only be gathered from new, independent empirical observation. This was Gauss's aim, according to Hilbert, when he confirmed the theorem for the first time, by measuring the angles of the large triangle formed by the three mountain peaks.[61] The network of concepts that constitute geometry, Hilbert concluded, has been proved consistent, and therefore it exists mathematically, independently of any observation. Whether or not

59 As in many other places in his lectures, Hilbert gave no direct reference to the specific physical theory he had in mind here, and in this particular case I have not been able to find it.

60 "Ich schließe hier noch die Bemerkung an, daß man jedes solches Begriffschema, das wir so rein logisch aus irgend welchen Axiomen aufbauen, anwenden kann auf beliebige gegenständliche Dinge, wenn sie nur diesen Axiomen genügen. ... Ein solches Beispiel für die Anwendung des Begriffschemas der nichteuklidischen Geometrie bildet das System der Lichtwege in unserer Atmosphäre unter dem Einfluß deren variabler Dichte und Brechungsexponenten; machen wir nämlich die einfachste mögliche Annahme, daß die Lichtgeschwindigkeit proportional ist dem vertikalen Abstande y von einer Horizontalebene, so ergeben sich als Lichtwege gerade die Orthogonalkreise jener Ebene, als Lichtzeit gerade die nichteuklidiche Entfernung auf ihnen. Um die hier obwaltenden Verhältnisse also genauer zu untersuchen, können wir gerade mit Vorteil das Begriffschema der nichteuklidischen Geometrie anwenden." (Hilbert 1905a, 69–70)

61 "In diesem Sinne und zu diesem Zwecke hat zuerst Gauß durch Messung an großen Dreiecken den Satz bestätigt." (Hilbert 1905a, 98)

it corresponds to reality is a question that can be decided only by observation, and our analysis of the independence of the axioms allows determining very precisely the minimal set of observations needed in order to do so.[62] Later on, he added, the same kind of perspective must be adopted concerning physical theories, although there its application will turn out to be much more difficult than in geometry.

In concluding his treatment of geometry, and before proceeding to discuss the specific axiomatization of individual physical theories, Hilbert summarized the role of the axiomatic method in a passage which encapsulates his view of science and of mathematics as living organisms whose development involves both an expansion in scope and an ongoing clarification of the logical structure of their existing parts.[63] The axiomatic treatment of a discipline concerns the latter; it is an important part of this growth but—Hilbert emphasized—only one part of it. The passage, reads as follows:

> The edifice of science is not raised like a dwelling, in which the foundations are first firmly laid and only then one proceeds to construct and to enlarge the rooms. Science prefers to secure as soon as possible comfortable spaces to wander around and only subsequently, when signs appear here and there that the loose foundations are not able to sustain the expansion of the rooms, it sets about supporting and fortifying them. This is not a weakness, but rather the right and healthy path of development.[64]

This metaphor provides the ideal background for understanding what Hilbert went on to realize at this point in his lectures, namely, to present his first detailed account of how the general idea of axiomatization of physical theories would be actually implemented in each case. But before we can really discuss that detailed account, it is necessary to broaden its context by briefly describing some relevant developments in physics just before 1905, and how they were manifest in Göttingen.

5. HILBERT AND PHYSICS IN GÖTTINGEN CIRCA 1905

The previous section described Hilbert's foundational activities in mathematics between 1900 and 1905. Those activities constituted the natural outgrowth of the seeds planted in *Grundlagen der Geometrie* and the developments that immediately

---

62 "Das Begriffsfachwerk der Geometrie selbst ist nach Erweisung seiner Widerspruchslosigkeit natürlich auch unabhängig von jeder Beobachtung mathematisch existent; der Nachweis seiner Übereinstimmung mit der Wirklichkeit kann nur durch Beobachtungen geführt werden, und die kleinste notwendige solche wird durch die Unabhängigkeitsuntersuchungen gegeben." (Hilbert 1905a, 98)

63 Elsewhere Hilbert called these two aspects of mathematics the "progressive" and "regressive" functions of mathematics, respectively (both terms not intended as value judgments, of course). See (Hilbert 1992, 17–18).

64 "Das Gebäude der Wissenschaft wird nicht aufgerichtet wie ein Wohnhaus, wo zuerst die Grundmauern fest fundiert werden und man dann erst zum Auf- und Ausbau der Wohnräume schreitet; die Wissenschaft zieht es vor, sich möglichst schnell wohnliche Räume zu verschaffen, in denen sie schalten kann, und erst nachträglich, wenn es sich zeigt, dass hier und da die locker gefügten Fundamente den Ausbau der Wohnräume nicht zu tragen vermögen, geht sie daran, dieselben zu stützen und zu befestigen. Das ist kein Mangel, sondern die richtige und gesunde Entwicklung." (Hilbert 1905a, 102.) Other places where Hilbert uses a similar metaphor are (Hilbert 1897, 67; Hilbert 1918, 148).

followed it. My account is not meant to imply, however, that Hilbert's focus of interest was limited to, or even particularly focused around, this kind of enquiry during those years. On 18 September 1901, for instance, Hilbert's keynote address at the commemoration of the 150th anniversary of the Göttingen Scientific Society (*Gesellschaft der Wissenschaften zu Göttingen*) was devoted to analyzing the conditions of validity of the Dirichlet principle (Hilbert 1904, 1905d). Although thus far he had published very little in this field, Hilbert's best efforts from then on would be given to analysis, and in particular, the theory of linear integral equations. His first publication in this field appeared in 1902, and others followed, up until 1912. But at the same time, he sustained his interest in physics, which is directly connected with analysis and related fields to begin with, and this interest in physics became only more diverse throughout this period. His increased interest in analysis had a natural affinity with the courses on potential theory (winter semester, 1901–1902; summer semester, 1902) and on continuum mechanics (winter semester, 1902–1903; summer semester, 1903) that he taught at that time. Perhaps worthy of greater attention, however, is Hilbert's systematic involvement around 1905 with the theories of the electron, on which I will elaborate in the present section.

Still, a brief remark on Hilbert's courses on continuum mechanics: The lecture notes of these two semesters (Hilbert 1902–1903, 1903b) are remarkable for the thoroughness with which the subject was surveyed. The presentation was probably the most systematic and detailed among all physical topics taught by Hilbert so far, and it comprised detailed examinations of the various existing approaches (particularly those of Lagrange, Euler and Helmholtz). Back in 1898–1899, in the final part of a course on mechanics, Hilbert had briefly dealt with the mechanics of systems of an infinite number of mass-points while stressing that the detailed analysis of such systems would actually belong to a different part of physics. This was precisely the subject he would consider in 1902. In that earlier course Hilbert had also discussed some variational principles of mechanics, without however presenting the theory in anything like a truly axiomatic perspective. Soon thereafter, in 1900 in Paris, Hilbert publicly presented his call for the axiomatization of physics. But in 1902–1903, in spite of the high level of detail with which he systematically discussed the physical discipline of continuum mechanics, the axiomatic presentation was not yet the guiding principle. Hilbert did state that a main task to be pursued was the axiomatic description of physical theories[65] and throughout the text he specifically accorded the status of axioms to some central statements.[66] Still, the notes convey the distinct impression that Hilbert did not believe that the time was ripe for the fully axiomatic

65 The manuscript shows an interesting hesitation on how this claim was stated: "Das <Als ein wichtiges> Ziel der Vorlesung ist <denke ich mir> die mathematische Beschreibung der Axiome der Physik. Vergl. Archiv der Mathematik und Physik, meine Rede: 'Probleme der Mathematik'." However, it is not clear if this amendment of the text reflects a hesitation on the side of Hilbert, or on the side of Berkowski, who wrote down the notes. (Hilbert 1902–1903, 2)

66 Thus for instance in (Hilbert 1902–1903).

treatment of mechanics, or at least of continuum mechanics, in axiomatic terms similar to those previously deployed in full for geometry.

On the other hand, it is worth stressing that in many places Hilbert set out to develop a possible unified conception of mechanics, thermodynamics (Hilbert 1903b, 47–91) and electrodynamics (Hilbert 1903b, 91–164) by using formal analogies with the underlying ideas of his presentation of the mechanics of continua. These ideas, which were treated in greater detail from an axiomatic point of view in the 1905 lectures, are described more fully below; therefore, at this point I will not give a complete account of them. Suffice it to say that Hilbert considered the material in these courses to be original and important, and not merely a simple repetition of existing presentations. In fact, the only two talks he delivered in 1903 at the meetings of the *Göttinger Mathematische Gesellschaft* were dedicated to reporting on their contents.[67]

Still in 1903, Hilbert gave a joint seminar with Minkowski on stability theory.[68] He also presented a lecture on the same topic at the yearly meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte* at Kassel,[69] sparking a lively discussion with Boltzmann.[70] In the winter semester of 1904–1905 Hilbert taught an exercise course on mechanics and later gave a seminar on the same topic. The course "Logical Principles of Mathematical Thinking," containing the lectures on axiomatization of physics, was taught in the summer semester of 1905. He then lectured again on mechanics (winter semester, 1905–1906) and two additional semesters on continuum mechanics.

The renewed encounter with Minkowski signified a major source of intellectual stimulation for these two old friends, and it particularly offered a noteworthy impulse to the expansion of Hilbert's horizon in physics. As usual, their walks were an opportunity to discuss a wide variety of mathematical topics, but now physics became a more prominent, common interest than it had been in the past. Teaching in Zürich since 1894, Minkowski had kept alive his interest in mathematical physics, and in particular in analytical mechanics and thermodynamics (Rüdenberg and Zassenhaus 1973, 110–114). Now at Göttingen, he further developed these interests. In 1906 Minkowski published an article on capillarity (Minkowski 1906), commissioned for

67 See the announcements in *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12 (1903), 226 and 445. Earlier volumes of the *Jahresbericht der Deutschen Mathematiker-Vereinigung* do not contain announcements of the activities of the *Göttinger Mathematische Gesellschaft*, and therefore it is not known whether he also discussed his earlier courses there.

68 *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 570/1) contains a somewhat random collection of handwritten notes related to many different courses and seminars of Hilbert. Notes of this seminar on stability theory appear on pp. 18–24. Additional, related notes appear in (Cod. Ms. D. Hilbert, 696).

69 *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 593) contains what appear to be the handwritten notes of this talk, with the title "Vortrag über Stabilität einer Flüssigkeit in einem Gefässe," and includes some related bibliography.

70 As reported in *Naturwissenschaftliche Rundschau*, vol. 18, (1903), 553–556 (cf. Schirrmacher 2003, 318, note 63). The reporter of this meeting, however, considered that Hilbert was addressing a subtlety, rather than a truly important physical problem.

the physics volume of the *Encyklopädie*, edited by Sommerfeld. At several meetings of the *Göttinger Mathematische Gesellschaft*, Minkowski lectured on this as well as other physical issues, such as Euler's equations of hydrodynamics and recent work on thermodynamics by Walter Nernst (1864–1941), (Nernst 1906), who by that time had already left Göttingen. Minkowski also taught advanced seminars on physical topics and more basic courses on mechanics, continuum mechanics, and exercises on mechanics and heat radiation.[71] In 1905 Hilbert and Minkowski organized, together with other Göttingen professors, an advanced seminar that studied recent progress in the theories of the electron.[72] In December 1906, Minkowski reported to the *Göttinger Mathematische Gesellschaft* on recent developments in radiation theory, and discussed the works of Hendrik Antoon Lorentz (1853–1928), Max Planck (1858–1947), Wilhelm Wien (1864–1928) and Lord Rayleigh (1842–1919), (Minkowski 1907, 78). Yet again in 1907, the two conducted a joint seminar on the equations of electrodynamics, and that semester Minkowski taught a course on heat radiation, after having studied with Hilbert Planck's recent book on this topic (Planck 1906).[73] Finally, as it is well known, during the last years of his life, 1907 to 1909, Minkowski's efforts were intensively dedicated to electrodynamics and the principle of relativity.

The 1905 electron theory seminar exemplifies the kind of unique scientific event that could be staged only at Göttingen at that time, in which leading mathematicians and physicists would meet on a weekly basis in order to intensively discuss current open issues of the discipline. In fact, over the preceding few years the *Göttinger Mathematische Gesellschaft* had dedicated many of its regular meetings to discussing recent works on electron theory and related topics, so that this seminar was a natural continuation of a more sustained, general interest for the local scientific community.

---

71 Cf. *Jahresbericht der Deutschen Mathematikervereinigung* 13 (1904), 492; 16 (1907), 171; 17 (1908), 116. See also the *Vorlesungsverzeichnisse*, Universität Göttingen, winter semester, 1903–1904, 14; summer semester, 1904, 14–16. A relatively large collection of documents and manuscripts from Minkowski's *Nachlass* has recently been made available at the Jewish National Library, at the Hebrew University, Jerusalem. These documents are yet to be thoroughly studied and analyzed. They contain scattered notes of courses taught at Königsberg, Bonn, Zurich and Göttingen. The notes of a Göttingen course on mechanics, winter semester, 1903–1904, are found in Box IX (folder 4) of that collection. One noteworthy aspect of these notes is that Minkowski's recommended reading list is very similar to that of Hilbert's earlier courses and comprises mainly texts then available at the *Lesezimmer*. It included classics such as Lagrange, Kirchhoff, Helmholtz, Mach, and Thomson-Tait, together with more recent, standard items such as the textbooks by Voigt, Appell, Petersen, Budde and Routh. Like Hilbert's list it also included the lesser known (Rausenberg 1888), but it also comprised two items absent from Hilbert's list: (Duhamel 1853–1854) and (Föppl 1901). Further, it recommended Voss's *Encyklopädie* article as a good summary of the field.

72 Pyenson (1979) contains a detailed and painstaking reconstruction of the ideas discussed in this seminar and the contributions of its participants. This reconstruction is based mainly on *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 570/9). I strongly relied on this article as a starting point for my account of the seminar in the next several paragraphs. Still, my account departs from Pyenson's views in some respects.

73 The notes of the course appear in (Minkowski 1907).

Besides Minkowski and Hilbert, the seminar was led by Wiechert and Gustav Herglotz (1881–1953). Herglotz had recently joined the Göttingen faculty and received his Habilitation for mathematics and astronomy in 1904. Alongside Wiechert, he contributed important new ideas to the electron theory and the two would later become the leading geophysicists of their time. The list of students who attended the seminar includes, in retrospect, no less impressive names: two future Nobel laureates, Max von Laue (1879–1960) and Max Born (1882–1970), as well as Paul Heinrich Blasius (1883–1970) who would later distinguish himself in fluid mechanics, and Arnold Kohlschütter (1883–1969), a student of Schwarzschild who became a leading astronomer himself. Parallel to this seminar, a second one on electrotechnology was held with the participation of Felix Klein, Carl Runge (1856–1914), Ludwig Prandtl (1875–1953) and Hermann Theodor Simon (1870–1918), then head of the Göttingen Institute for Applied Electricity.[74]

The modern theory of the electron had developed through the 1890s, primarily with the contributions of Lorentz working in Leiden, but also through the efforts of Wiechert at Göttingen and—following a somewhat different outlook—of Joseph Larmor (1857–1942) at Cambridge.[75] Lorentz had attempted to account for the interaction between aether and matter in terms of rigid, negatively charged, particles: the electrons. His article of 1895 dealing with concepts such as stationary aether and local time, while postulating the existence of electrons, became especially influential (Lorentz 1895). The views embodied in Lorentz's and Larmor's theories received further impetus from contemporary experimental work, such as that of Pieter Zeeman (1865–1943) on the effect associated with his name, work by J. J. Thomson (1856–1940) especially concerning the cathode ray phenomena and their interpretation in terms of particles, and also work by Wiechert himself, Wien and Walter Kaufmann (1871–1947). Gradually, the particles postulated by the theories and the particle-laden explanations stemming from the experiments came to be identified with one another.[76]

Lorentz's theory comprised elements from both Newtonian mechanics and Maxwell's electrodynamics. While the properties of matter are governed by Newton's laws, Maxwell's equations describe the electric and magnetic fields, conceived as states of the stationary aether. The electron postulated by the theory provided the connecting link between matter and aether. Electrons moving in the aether generate electric and magnetic fields, and the latter exert forces on material bodies through the electrons themselves. The fact that Newton's laws are invariant under Galilean transformations and Maxwell's are invariant under what came to be known as Lorentz transformations was from the outset a source of potential problems and difficulties for the theory, and in a sense, these and other difficulties would be dispelled only with the formulation of Einstein's special theory of relativity in 1905. In Lorentz's theory

74 Cf. (Pyenson 1979, 102).
75 Cf. (Warwick 1991).
76 Cf. (Arabatzis 1996).

the conflict with experimental evidence led to the introduction of the famous contraction hypothesis and in fact, of a deformable electron.[77] But in addition it turned out that, in this theory, some of the laws governing the behavior of matter would be Lorentz invariant, rather than Galilean, invariant. The question thus arose whether this formal, common underlying property does not actually indicate a more essential affinity between what seemed to be separate realms, and, in fact, whether it would not be possible to reduce all physical phenomena to electrodynamics.[78]

Initially, Lorentz himself attempted to expand the scope of his theory, as a possible foundational perspective for the whole of physics, and in particular as a way to explain molecular forces in terms of electrical ones. He very soon foresaw a major difficulty in subsuming also gravitation within this explanatory scope. Still, he believed that such a difficulty could be overcome, and in 1900 he actually published a possible account of gravitation in terms of his theory. The main difficulty in this explanation was that, according to existing astronomical data, the velocity of gravitational effects would seem to have to expand much faster than electromagnetic ones, contrary to the requirements of the theory (Lorentz 1900). This and other related difficulties are in the background of Lorentz's gradual abandonment of a more committed foundational stance in connection with electron theory and the electromagnetic worldview. But the approach he had suggested in order to address gravitational phenomena in electromagnetic terms was taken over and further developed that same year by Wilhelm Wien, who had a much wider aim. Wien explicitly declared that his goal was to unify currently "isolated areas of mechanical and electromagnetic phenomena," and in fact, to do so in terms of the theory of the electron while assuming that all mass was electromagnetic in nature, and that Newton's laws of mechanics should be reinterpreted in electromagnetic terms.[79]

One particular event that highlighted the centrality of the study of the connection and interaction between aether and matter in motion among physicists in the German-speaking world was the 1898 meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte*, held at Düsseldorf jointly with the annual meeting of the *Deutsche Mathematiker- Vereinigung*. Most likely both Hilbert and Minkowski had the opportunity to attend Lorentz's talk, which was the focus of interest. Lorentz described the main problem facing current research in electrodynamics in the following terms:

> Ether, ponderable matter, and, we may add, electricity are the building stones from which we compose the material world, and if we could know whether matter, when it moves, carries the ether with it or not, then the way would be opened before us by which

77 In Larmor's theory the situation was slightly different, and so were the theoretical reasons for adopting the contraction hypothesis, due also to George FitzGerald (1851–1901). For details, see (Warwick 2003, 367–376).

78 For a more detailed explanation, cf. (Janssen 2002).

79 See (Wien 1900). This is the article to which Voss referred in his survey of 1901, and that he took to be representative of the new foundationalist trends in physics. Cf. (Jungnickel and McCormmach 1986, 2: 236–240).

> we could further penetrate into the nature of these building stones and their mutual relations. (Lorentz 1898, 101)[80]

This formulation was to surface again in Hilbert's and Minkowski's lectures and seminars on electrodynamics after 1905.

The theory of the electron itself was significantly developed in Göttingen after 1900, with contributions to both its experimental and theoretical aspects. The experimental side came up in the work of Walter Kaufmann, who had arrived from Berlin in 1899. Kaufmann experimented with Becquerel rays, which produced high-speed electrons. Lorentz's theory stipulated a dependence of the mass of the electron on its velocity v, in terms of a second order relation on v/c (c being, of course, the speed of light). In order to confirm this relation it was necessary to observe electrons moving at speeds as close as possible to c, and this was precisely what Kaufmann's experiments could afford, by measuring the deflection of electrons in electric and magnetic fields. He was confident of the possibility of a purely electromagnetic physics, including the solution of open issues such as the apparent character of mass, and the gravitation theory of the electron. In 1902 he claimed that his results, combined with the recent developments of the theory, had definitely confirmed that the mass of the electrons is of "purely electromagnetic nature."[81]

The recent developments of the theory referred to by Kaufmann were those of his colleague at Göttingen, the brilliant *Privatdozent* Max Abraham (1875–1922). In a series of publications, Abraham introduced concepts such as "transverse inertia," and "longitudinal mass," on the basis of which he explained where the dynamics of the electron differed from that of macroscopic bodies. He also developed the idea of a rigid electron, as opposed to Lorentz's deformable one. He argued that explaining the deformation of the electron as required in Lorentz's theory would imply the need to introduce inner forces of non-electromagnetic origin, thus contradicting the most fundamental idea of a purely electromagnetic worldview. In Abraham's theory, the kinematic equations of a rigid body become fundamental, and he introduced variational principles to derive them. Thus, for instance, using a Lagrangian equal to the difference between the magnetic and the electrical energy, Abraham described the translational motion of the electron and showed that the principle of least action also holds for what he called "quasi-stationary" translational motion (namely, motion in which the velocity of the electron undergoes a small variation over the time required for light to traverse its diameter). Abraham attributed special epistemological significance to the fact that the dynamics of the electron could be expressed by means of a Lagrangian (Abraham 1903, 168),[82] a point that will surface interestingly in Hilbert's 1905 lectures on axiomatization, as we will see in the next section. Beyond the technical level, Abraham was a staunch promoter of the electromagnetic worldview and his theory of the electron was explicitly conceived to "shake the foundations of the

80 Translation quoted from (Hirosige 1976, 35).

81 Cf. (Hon 1995; Miller 1997, 44–51, 57–62).

82 On Abraham's electron theory, see (Goldberg 1970; Miller 1997, 51–57).

mechanical view of nature." Still, in 1905 he conceded that "the electromagnetic world picture is so far only a program."[83]

Among the organizers of the 1905 electron theory seminar, it was Wiechert who had been more directly involved in research of closely related issues. Early in his career he became fascinated by the unification of optics and electromagnetism offered by Maxwell's theory, and was convinced of the centrality of the aether for explaining all physical phenomena. In the 1890s, still unaware of Lorentz's work, he published the outlines of his own theory of "atoms of electricity," a theory which he judged to be still rather hypothetical, however. This work contained interesting theoretical and experimental aspects that supported his view that cathode ray particles were indeed the electric atoms of his theory. After his arrival in Göttingen in 1897, Wiechert learnt about Lorentz's theory, and quickly acknowledged the latter's priority in developing an electrodynamics based on the concept of the "electron," the term that he now also adopted. Like Lorentz, Wiechert also adopted a less committed and more skeptical approach towards the possibility of a purely electromagnetic foundation of physics.[84] Obviously Hilbert was in close, continued contact with Wiechert and his ideas, but one rather remarkable opportunity to inspect these ideas more closely came up once again in 1899, when Wiechert published an article on the foundations of electrodynamics as the second half of the *Gauss-Weber Festschrift* (Wiechert 1899).

Not surprisingly, Abraham's works on electron theory were accorded particular attention by his Göttingen colleagues in the 1905 seminar, yet Abraham himself seems not to have attended the meetings in person. He was infamous for his extremely antagonistic and aggressive personality,[85] and this background may partly explain his absence. But one wonders if also his insistence on the foundational implications of electron theory, and a completely different attitude of the seminar leaders to this question may provide an additional, partial explanation for this odd situation. I already mentioned Wiechert's basic skepticism, or at least caution, in this regard. As we will see, also Hilbert and Minkowski were far from wholeheartedly supporting a purely electromagnetic worldview. Kaufmann was closest to Abraham in this point, and he had anyway left Göttingen in 1903. It is interesting to notice, at any rate, that Göttingen physicists and mathematicians held different, and very often conflicting, views on this as well as other basic issues, and it would be misleading to speak of a "Göttingen approach" to any specific topic. The situation around the electron theory seminar sheds interesting light on this fact.

Be that as it may, the organizers relied not on Abraham's, but on other, different works as the seminar's main texts. The texts included, in the first place, Lorentz's 1895 presentation of the theory, and also his more recently published *Encyklopädie*

83 Quoted in (Jungnickel and McCormmach 1986, 2: 241). For a recent summary account of the electromagnetic worldview and the fate of its program, see (Kragh 1999, 105–199).

84 Cf. (Darrigol 2000, 344–347).

85 Cf., e.g., (Born 1978, 91 and 134–137).

article (Lorentz 1904a), which was to become the standard reference in the field for many years to come. Like most other surveys published in the *Encyklopädie*, Lorentz's article presented an exhaustive and systematic examination of the known results and existing literature in the field, including the most recent. The third basic text used in the seminar was Poincaré's treatise on electricity and optics (Poincaré 1901), based on his Sorbonne lectures of 1888, 1890 and 1891. This text discussed the various existing theories of the electrodynamics of moving bodies and criticized certain aspects of Lorentz's theory, and in particular a possible violation of the reaction principle due to its separation of matter and aether.[86]

Alongside the texts of Lorentz, Poincaré and Abraham, additional relevant works by Göttingen scientists were also studied. In fact, the main ideas of Abraham's theory had been recently elaborated by Schwarzschild and by Paul Hertz (1881–1940). The latter wrote a doctoral dissertation under the effective direction of Abraham, and this dissertation was studied at the seminar together with Schwarzschild's paper (Hertz 1904; Schwarzschild 1903). So were several recent papers by Sommerfeld (1904a, 1904b, 1905) who was now at Aachen, but who kept his strong ties to Göttingen always alive. Naturally, the ideas presented in the relevant works of Herglotz and Wiechert were also studied in the seminar (Herglotz 1903; Wiechert 1901).

The participants in this seminar discussed the current state of the theory, the relevant experimental work connected with it, and some of its most pressing open problems. The latter included the nature of the mass of the electron, problems related to rotation, vibration and acceleration in electron motion and their effects on the electromagnetic field, and the problem of faster-than-light motion. More briefly, they also studied the theory of dispersion and the Zeeman effect. From the point of view of the immediate development of the theory of relativity, it is indeed puzzling, as Lewis Pyenson has rightly stressed in his study of the seminar, that the participants were nowhere close to achieving the surprising breakthrough that Albert Einstein (1879–1956) had achieved at roughly the same time, and was about to publish (Pyenson 1979, 129–131).[87] Nevertheless, from the broader point of view of the development of math-

86 Cf. (Darrigol 2000, 351–366).

87 According to Pyenson, whereas Einstein "sought above all to address the most essential properties of nature," the Göttingen seminarists "sought to subdue nature, as it were, by the use of pure mathematics. They were not much interested in calculating with experimentally observable phenomena. They avoided studying electrons in metal conductors or at very low or high temperatures, and they did not spend much time elaborating the role of electrons in atomic spectra, a field of experimental physics then attracting the interest of scores of young physicists in their doctoral dissertations." Pyenson stresses the fact that Ritz's experiment was totally ignored at the seminar and adds: "For the seminar Dozenten it did not matter that accelerating an electron to velocities greater than that of light and even to infinite velocities made little physical sense. They pursued the problem because of its intrinsic, abstract interest." Noteworthy as these points are, it seems to me that by overstressing the question of why the Göttingen group achieved less than Einstein did, the main point is obscured in Pyenson's article, namely, what and why were Hilbert, Minkowski and their friends doing what they were doing, and how is this connected to the broader picture of their individual works and of the whole Göttingen mathematical culture.

ematics and physics at the turn of the century, and particularly of the account pursued here, it is all the more surprising to notice the level of detail and close acquaintance with physical theory and also, to a lesser degree, with experiment, that mathematicians such as Hilbert and Minkowski had attained by that time. All this, of course, while they were simultaneously active and highly productive in their own main fields of current, purely mathematical investigations: analysis, number theory, foundations, etc. Hilbert's involvement in the electron theory seminar clarifies the breadth and depth of the physical background that underlie his lectures on the axiomatization of physics in 1905, and that had considerably expanded in comparison with the one that prompted him to formulate, in the first place, his sixth problem back in 1900.

6. AXIOMS FOR PHYSICAL THEORIES: HILBERT'S 1905 LECTURES

Having described in some detail the relevant background, I now proceed to examine the contents of Hilbert's 1905 lectures on the "Axiomatization of Physical Theories," which devote separate sections to the following topics:

- Mechanics
- Thermodynamics
- Probability Calculus
- Kinetic Theory of Gases
- Insurance Mathematics
- Electrodynamics
- Psychophysics

Here I shall limit myself to discussing the sections on mechanics, the kinetic theory of gases, and electrodynamics.

6.1 Mechanics

Clearly, the main source of inspiration for this section is Aurel Voss's 1901 *Encyklopädie* article (Voss 1901). This is evident in the topics discussed, the authors quoted, the characterization of the possible kinds of axioms for physics, the specific axioms discussed, and sometimes even the order of exposition. Hilbert does not copy Voss, of course, and he introduces many ideas and formulations of his own, and yet the influence is clearly recognizable.

Though at this time Hilbert considered the axiomatization of physics and of natural science in general to be a task whose realization was still very distant,[88] we can mention one particular topic for which the axiomatic treatment had been almost com-

88 "Von einer durchgeführten axiomatischen Behandlung der Physik und der Naturwissenschaften ist man noch weit entfernt; nur auf einzelnen Teilgebieten finden sich Ansätze dazu, die nur in ganz wenigen Fällen durchgeführt sind. <Die Durchführung ist ein ganzes—grosses—Arbeitsprogramm, Vgl. Dissertation von Schimmack sowie Schur>." (Hilbert 1905a, 121)

pletely attained (and only very recently, for that matter). This is the "law of the parallelogram" or, what amounts to the same thing, the laws of vector-addition. Hilbert based his own axiomatic presentation of this topic on works by Darboux, by Hamel, and by one of his own students, Rudolf Schimmack (1881–1912).[89]

Hilbert defined a force as a three-component vector, and made no additional, explicit assumptions here about the nature of the vectors themselves, but it is implicitly clear that he had in mind the collection of all ordered triples of real numbers. Thus, as in his axiomatization of geometry, Hilbert was not referring to an arbitrary collection of abstract objects, but to a very concrete mathematical entity; in this case, one that had been increasingly adopted after 1890 in the treatment of physical theories, following the work of Oliver Heaviside (1850–1925) and Josiah Willard Gibbs (1839–1903).[90] In fact, in Schimmack's article of 1903—based on his doctoral dissertation—a vector was explicitly defined as a directed, real segment of line in the Euclidean space. Moreover, Schimmack defined two vectors as equal when their lengths as well as their directions coincide (Schimmack 1903, 318).

The axioms presented here were thus meant to define the addition of two such given vectors, as the sums of the components of the given vectors. At first sight, this very formulation could be taken as the single axiom needed to define the sum. But the task of axiomatic analysis is precisely to separate this single idea into a system of several, mutually independent, simpler notions that express the basic intuitions involved in it. Otherwise, it would be like taking the linearity of the equation representing the straight line as the starting point of geometry.[91] Hilbert had shown in his previous discussion on geometry that this latter result could be derived using all his axioms of geometry.

Hilbert thus formulated six axioms to define the addition of vectors: the first three assert the existence of a well-defined sum for any two given vectors (without stating what its value is), and the commutativity and associativity of this operation. The fourth axiom connects the resultant vector with the directions of the summed vectors as follows:

> 4. Let aA denote the vector $(aAx, aAy$,aAz $)$, having the same direction as A. Then every real number a defines the sum:
>
> $$A + aA = (1 + a)A.$$
>
> i.e., the addition of two vectors having the same direction is defined as the algebraic addition of the extensions along the straight line on which both vectors lie.[92]

89 The works referred to by Hilbert are (Darboux 1875; Hamel 1905; Schimmack 1903). Schimmak's paper was presented to the *Königliche Gesellschaft der Wissenschaften zu Göttingen* by Hilbert himself. An additional related work, also mentioned by Hilbert in the manuscript, is (Schur 1903).

90 Cf. (Crowe 1967, 150 ff.; Yavetz 1995).

91 "... das andere wäre genau dasselbe, wie wenn man in der Geometrie die Linearität der Geraden als einziges Axiom an die Spitze stellen wollte (vgl. S. 118)." (Hilbert 1905a, 123)

92 "Addition zweier Vektoren derselben Richtung geschieht durch algebraische Addition der Strecken auf der gemeinsamen Geraden." (Hilbert 1905a, 123)

The fifth one connects addition with rotation:

> 5. If D denotes a rotation of space around the common origin of two forces A and B, then the rotation of the sum of the vectors equals the sum of the two rotated vectors:
>
> $$D(A+B) = DA + DB$$
>
> i.e., the relative position of sum and components is invariant with respect to rotation.[93]

The sixth axiom concerns continuity:

> 6. Addition is a continuous operation, i.e., given a sufficiently small domain G around the end-point of $A+B$ one can always find domains G_1 and G_2, around the endpoints of A and B respectively, such that the endpoint of the sum of any two vectors belonging to each of these domains will always fall inside G.[94]

These are all simple axioms—Hilbert continued, without having really explained what a "simple" axiom is—and if we think of the vectors as representing forces, they also seem rather plausible. The axioms thus correspond to the basic known facts of experience, i.e., that the action of two forces on a point may always be replaced by a single one; that the order and the way in which they are added do not change the result; that two forces having one and the same direction can be replaced by a single force having the same direction; and, finally, that the relative position of the components and the resultant is independent of rotations of the coordinates. Finally, the demand for continuity in this system is similar and is formulated similarly to that of geometry.

That these six axioms are in fact necessary to define the law of the parallelogram was first claimed by Darboux, and later proven by Hamel. The main difficulties for this proof arose from the sixth axiom. Schimmack proved in 1903 the independence of the six axioms (in a somewhat different formulation), using the usual technique of models that satisfy all but one of the axioms. Hilbert also mentioned some possible modifications of this system. Thus, Darboux himself had showed that the continuity axiom may be abandoned, and in its place, it may be postulated that the resultant lies on the same plane as, and within the internal angle between, the two added vectors. Hamel, on the other hand, following a conjecture of Friedrich Schur, proved that the fifth axiom is superfluous if we assume that the locations of the endpoints of the resultants, seen as functions of the two added vectors, have a continuous derivative. In fact—Hilbert concluded—if we assume that all functions appearing in the natural sciences have at least one continuous derivative, and take this assumption as an even

---

93 "Nimmt man eine Drehung D des Zahlenraumes um den gemeinsamen Anfangspunkt vor, so entsteht aus $A+B$ die Summe der aus A und B durch D entstehenden Vektoren: $D(A+B) = DA+DB$; d.h. die relative Lage von Summe und Komponenten ist gegenüber allen Drehungen invariant." (Hilbert 1905a, 124)

94 "Zu einem genügend kleinen Gebiete G um den Endpunkt von $A+B$ kann man stets um die Endpunkte von A und B solche Gebiete G_1, G_2 abgrenzen, daß der Endpunkt der Summe jedes in G_1 u. G_2 endigenden Vectorpaares nach G fällt." (Hilbert 1905a, 124)

more basic axiom, then vector addition is defined by reference to only the four first axioms in the system.[95]

The sixth axiom, the axiom of continuity, plays a very central role in Hilbert's overall conception of the axiomatization of natural science—geometry, of course, included. It is part of the essence of things—Hilbert said in his lecture—that the axiom of continuity should appear in every geometrical or physical system. Therefore it can be formulated not just with reference to a specific domain, as was the case here for vector addition, but in a much more general way. A very similar opinion had been advanced by Hertz, as we saw, who described continuity as "an experience of the most general kind," and who saw it as a very basic assumption of all physical science. Boltzmann, in his 1897 textbook, had also pointed out the continuity of motion as the first basic assumption of mechanics, which in turn should provide the basis for all of physical science.[96] Hilbert advanced in his lectures the following general formulation of the principle of continuity:

> If a sufficiently small degree of accuracy is prescribed in advance as our condition for the fulfillment of a certain statement, then an adequate domain may be determined, within which one can freely choose the arguments [of the function defining the statement], without however deviating from the statement, more than allowed by the prescribed degree.[97]

Experiment—Hilbert continued—compels us to place this axiom at the top of every natural science, since it allows us to assert the validity of our assumptions and claims.[98] In every special case, this general axiom must be given the appropriate version, as Hilbert had shown for geometry in an earlier part of the lectures and here for vector addition. Of course there are many important differences between the Archimedean axiom, and the one formulated here for physical theories, but Hilbert seems to have preferred stressing the similarity rather than sharpening these differences. In fact, he suggested that from a strictly mathematical point of view, it would be possible to conceive interesting systems of physical axioms that do without continuity, that is, axioms that define a kind of "non-Archimedean physics." He did not consider such systems here, however, since the task was to see how the ideas and methods of axiomatics could be fruitfully applied to physics.[99] Nevertheless, this is an extremely important topic in Hilbert's axiomatic treatment of physical theories. When speaking of applying axiomatic ideas and methods to these theories, Hilbert

---

95 "Nimmt man nun von vornherein als Grundaxiom aller Naturwissenschaft an, daß alle auftretenden Funktionen einmal stetig differenzierbar sind, so kommt man hier mit den ersten 4 Axiomen aus." (Hilbert 1905a, 127)

96 Quoted in (Boltzmann 1974, 228–229).

97 "Schreibt man für die Erfüllung der Behauptung einen gewissen genügend kleinen Genaugikeitsgrad vor, so läßt sich ein Bereich angeben, innerhalb dessen man die Voraussetzungen frei wählen kann, ohne daß die Abweichung der Behauptung jenen vorgeschriebenen Grad überschreitet." (Hilbert 1905a, 125)

98 "Das Experiment zwingt uns geradezu dazu, ein solches Axiom an die Spitze aller Naturwissenschaft zu setzen, denn wir können bei ihm stets nur das ~~Ein~~<Zu>treffen von Voraussetzung und Behauptung mit einer gewissen beschränkten Genauigkeit feststellen." (Hilbert 1905a, 125–126)

meant in this case existing physical theories. But the possibility suggested here, of examining models of theories that preserve the basic logical structure of classical physics, except for a particular feature, opens the way to the introduction and systematic analysis of alternative theories, close enough to the existing ones in relevant respects. Hilbert's future works on physics, and in particular his work on general relativity, would rely on the actualization of this possibility.

An additional point that should be stressed in relation to Hilbert's treatment of vector addition has to do with his disciplinary conceptions. The idea of a vector space, and the operations with vectors as part of it, has been considered an integral part of algebra at least since the 1920s.[100] This was not the case for Hilbert, who did not bother here to make any connection between his axioms for vector addition and, say, the already well-known axiomatic definition of an abstract group. For Hilbert, as for the other mathematicians he cites in this section, this topic was part of physics rather than of algebra.[101] In fact, the articles by Hamel and by Schur were published in the *Zeitschrift für Mathematik und Physik*—a journal that bore the explicit subtitle: *Organ für angewandte Mathematik*. It had been established by Oscar Xavier Schlömilch (1823–1901) and by the turn of the century its editor was Carl Runge, the leading applied mathematician at Göttingen.

After the addition of vectors, Hilbert went on to discuss a second domain related to mechanics: statics. Specifically, he considered the axioms that describe the equilibrium conditions of a rigid body. The main concept here is that of a force, which can be described as a vector with an application point. The state of equilibrium is defined by the following axioms:

> I. Forces with a common application point are equivalent to their sum.
>
> II. Given two forces, K, L, with different application points, P, Q, if they have the same direction, and the latter coincides with the straight line connecting P and Q, then these forces are equivalent.
>
> III. A rigid body is in a state of equilibrium if all the forces applied to it taken together are equivalent to 0.[102]

---

99 "Rein mathematisch werden natürlich auch <physikalische> Axiomensysteme, die auf <diese> Stetigkeit Verzicht leisten, also eine 'nicht-Archimedische Physik' in erweitertem Sinne definieren, von hohem Interesse sein können; wir werden jedoch zunächst noch von ihrer Betrachtung absehen können, da es sich vorerst überhaupt nur darum handelt, die fruchtbaren Ideen und Methoden der Axiomatik in die Physik einzuführen." (Hilbert 1905a, 126)

100 Cf. (Dorier 1995; Moore 1995).

101 This point, which helps understanding Hilbert's conception of algebra, is discussed in detail in (Corry 2003, §3.4).

102 "1., Kräfte mit demselben Angriffspunkt sind ihrer Summe (im obigen Sinne) '*aequivalent*.' 2., 2 Kräfte K, L mit verschiedenen Angriffspunkten P, Q und dem gleichen (auch gleichgerichteten) Vektor, deren Richtung in die Verbindung P, Q fällt, heißen gleichfalls *aequivalent*. … Ein starrer Körper befindet sich im *Gleichgewicht*, wenn die an ihm angreifenden Kräfte zusammengenommen der Null aequivalent sind." (Hilbert 1905a, 127–128)

From these axioms, Hilbert asserted, the known formulae of equilibrium of forces lying on the same plane (e.g., for the case of a lever and or an inclined plane) can be deduced. As in the case of vector addition, Hilbert's main aim in formulating the axioms was to uncover the basic, empirical facts that underlie our perception of the phenomenon of equilibrium.

In the following lectures Hilbert analyzed in more detail the principles of mechanics and, in particular, the laws of motion. In order to study motion, one starts by assuming space and adds time to it. Since geometry provides the axiomatic study of space, the axiomatic study of motion will call for a similar analysis of time.

According to Hilbert, two basic properties define time: (1) its uniform passage and (2) its unidimensionality.[103] A consistent application of Hilbert's axiomatic approach to this characterization immediately leads to the question: Are these two independent facts given by intuition,[104] or are they derivable the one from the other? Since this question had very seldom been pursued, he said, one could only give a brief sketch of tentative answers to it. The unidimensionality of time is manifest in the fact, that, whereas to determine a point in space one needs three parameters, for time one needs only the single parameter t. This parameter t could obviously be transformed, by changing the marks that appear on our clocks,[105] which is perhaps impractical but certainly makes no logical difference. One can even take a discontinuous function for t, provided it is invertible and one-to-one,[106] though in general one does not want to deviate from the continuity principle, desirable for all the natural sciences. Hilbert's brief characterization of time would seem to allude to Carl Neumann's (Neumann 1870), in his attempt to reduce the principle of inertia into simpler ones.

Whereas time and space are alike in that, for both, arbitrarily large values of the parameters are materially inaccessible, a further basic difference between them is that time can be experimentally investigated in only one direction, namely, that of its increase.[107] While this limitation is closely connected to the unidimensionality of time,[108] the issue of the uniform passage of time is an experimental fact, which has to be deduced, according to Hilbert, from mechanics alone.[109] This claim was elaborated into a rather obscure discussion of the uniform passage for which, as usual, Hilbert gave no direct references, but which clearly harks back to Hertz's and Larmor's

---

103 "... ihr *gleichmäßiger Verlauf* und ihre *Eindimensionalität*." (Hilbert 1905a, 128)

104 "... anschauliche unabhängige Tatsachen." (Hilbert 1905a, 129)

105 "Es ist ohne weiteres klar, daß dieser Parameter t durch eine beliebige Funktion von sich ersetzt werden kann; das würde etwa nur auf eine andere Benennung der Ziffern der Uhr oder einen unregelmäßigen Gang des Zeigers hinauskommen." (Hilbert 1905a, 129)

106 One is reminded here of a similar explanation, though in a more general context, found in Hilbert's letter to Frege, on 29 December 1899. See (Gabriel et al. 1980, 41).

107 "~~Der~~ <Ein> wesentlicher Unterschied von Zeit und Raum ist nur der, daß wir in der Zeit nur in einem Sinne, dem des *wachsenden Parameters* experimentieren können, während Raum und Zeit darin übereinstimmen, daß uns *beliebig große Parameterwerte* unzugänglich sind." (Hilbert 1905a, 129)

108 Here Hilbert adds with his own handwriting (p. 130): <Astronomie! Wie wichtig wäre Beobachtungen in ferner Vergangenheit u. Zukunft!>."

109 "... eine experimentelle nur aus der Mechanik zu entnehmende Tatsache." (Hilbert 1905a, 130)

discussions and referred to by Voss as well, as mentioned earlier. I try to reproduce Hilbert's account here without really claiming to understand it. In short, Hilbert argued that if the flow of time were non-uniform then an essential difference between organic and inorganic matter would be reflected in the laws of mechanics, which is not actually the case. He drew attention to the fact that the differential expression $m \cdot d^2x/dt^2$ characterizes a specific physical situation only when it vanishes, namely, in the case of inertial motion. From a logical point of view, however, there is no apparent reason why the same situation might not be represented in terms of a more complicated expression, e.g., an expression of the form

$$m_1 \frac{d^2x}{dt^2} + m_2 \frac{dx}{dt}.$$

The magnitudes m_1 and m_2 may depend not only on time, but also on the kind of matter involved,[110] e.g., on whether organic or inorganic matter is involved. By means of a suitable change of variables, $t = t(\tau)$, this latter expression could in turn be transformed into $\mu \cdot d^2x/d\tau^2$, which would also depend on the kind of matter involved. Thus different kinds of substances would yield, under a suitable change of variables, different values of "time," values that nevertheless still satisfy the standard equations of mechanics. One could then use the most common kind of matter in order to measure time,[111] and when small variations of organic matter occurred along large changes in inorganic matter, clearly distinguishable non-uniformities in the passage of time would arise.[112] However, it is an intuitive (*anschauliche*) fact, indeed a mechanical axiom, Hilbert said, that the expression $m \cdot d^2x/dt^2$, always appears in the equations with *one and the same* parameter t, independently of the kind of substance involved, contrary to what the above discussion would seem to imply. This latter fact, to which Hilbert wanted to accord the status of axiom, is then the one that establishes the uniform character of the passage of time. Whatever the meaning and the validity of this strange argument, one source where Hilbert was likely to have found it is Aurel Voss's 1901 *Encyklopädie* article, which quotes, in this regard, similar passages of Larmor and Hertz.[113]

Following this analysis of the basic ideas behind the concept of time, Hilbert repeated the same kind of reasoning he had used in an earlier lecture concerning the role of continuity in physics. He suggested the possibility of elaborating a non-Galilean mechanics, i.e., a mechanics in which the measurement of time would depend on the kind of matter involved, in contrast to the characterization presented earlier in his lecture. This mechanics would, in most respects, be in accordance with

110 "... die m_1, m_2 von der Zeit, vor allem aber von dem Stoffe abhängig sein können." (Hilbert 1905a, 130)

111 "... der häufigste Stoff etwa kann dann zu Zeitmessungen verwandt werden." (Hilbert 1905a, 130–131)

112 "... für uns leicht große scheinbare Unstetigkeiten der Zeit auftreten." (Hilbert 1905a, 131)

113 See (Voss 1901, 14). Voss quoted (Larmor 1900, 288) and (Hertz 1894, 165).

the usual one, and thus one would be able to recognize which parts of mechanics depend essentially on the peculiar properties of time, and which parts do not. It is only in this way that the essence of the uniform passage of time can be elucidated, he thought, and one may thus at last understand the exact scope of the connection between this property and the other axioms of mechanics.

So much for the properties of space and time. Hilbert went on to discuss the properties of motion, while concentrating on a single material point. This is clearly the simplest case and therefore it is convenient for Hilbert's axiomatic analysis. However, it must be stressed that Hilbert was thereby distancing himself from Hertz's presentation of mechanics, in which the dynamics of single points is not contemplated. One of the axioms of statics formulated earlier in the course stated that a point is in equilibrium when the forces acting on it are equivalent to the null force. From this axiom, Hilbert derived the Newtonian law of motion:

$$m.\frac{d^2x}{dt^2} = X; m.\frac{d^2y}{dt^2} = Y; m.\frac{d^2z}{dt^2} = Z.$$

Newton himself, said Hilbert, had attempted to formulate a system of axioms for his mechanics, but his system was not very sharply elaborated and several objections could be raised against it. A detailed criticism, said Hilbert, was advanced by Mach in his *Mechanik*.[114]

The above axiom of motion holds for a free particle. If there are constraints, e.g., that the point be on a plane $f(x, y, z) = 0$ then one must introduce an additional axiom, namely, Gauss's principle of minimal constraint. Gauss's principle establishes that a particle in nature moves along the path that minimizes the following magnitude:

$$\frac{1}{m}\{(mx'' - X)^2 + (my'' - Y)^2 + (mz'' - Z)^2\} = \text{Minim.}$$

Here x'',y'', and z'' denote the components of the acceleration of the particle, and X,Y,Z the components of the moving force. Clearly, although Hilbert did not say it in his manuscript, if the particle is free from constraints, the above magnitude can actually become zero and we simply obtain the Newtonian law of motion. If there are constraints, however, the magnitude can still be minimized, thus yielding the motion of the particle.[115]

---

114 A detailed account of the kind of criticism advanced by Mach, and before him by Carl Neumann and Ludwig Lange, appears in (Barbour 1989, chap. 12).

115 For more detail on Gauss's principle, see (Lanczos 1962, 106–110). Interestingly, Lanczos points out that "Gauss was much attached to this principle because it represents a perfect physical analogy to the 'method of least squares' (discovered by him and independently by Legendre) in the adjustment of errors." Hilbert also discussed this latter method in subsequent lectures, but did not explicitly make any connection between Gauss's two contributions.

In his lectures, Hilbert explained in some detail how the Lagrangian equations of motion could be derived from this principle. But he also stressed that the Lagrangian equations could themselves be taken as axioms and set at the top of the whole of mechanics. In this case, the Newtonian and Galilean principles would no longer be considered as necessary assumptions of mechanics. Rather, they would be logical consequences of a distinct principle. Although this is a convenient approach that is often adopted by physicists, Hilbert remarked, it has the same kinds of disadvantages as deriving the whole of geometry from the demand of linearity for the equations of the straight line: many results can be derived from it, but it does not indicate what the *simplest* assumptions underlying the considered discipline may be. All the discussion up to this point, said Hilbert, concerns the simplest and oldest systems of axioms for the mechanics of systems of points. Beside them there is a long list of other possible systems of axioms for mechanics. The first of these is connected to the principle of conservation of energy, which Hilbert associated with the law of the impossibility of a *perpetuum mobile* and formulated as follows: "If a system is at rest and no forces are applied, then the system will remain at rest."[116]

Now the interesting question arises, how far can we develop the whole of mechanics by putting this law at the top? One should follow a process similar to the one applied in earlier lectures: to take a certain result that can be logically derived from the axioms and try to find out if, and to what extent, it can simply replace the basic axioms. In this case, it turns out that the law of conservation alone, as formulated above, is sufficient, though not necessary, for the derivation of the conditions of equilibrium in mechanics.[117] In order to account for the necessary conditions as well, the following axiom must be added: "A mechanical system can *only* be in equilibrium if, in accordance with the axiom of the impossibility of a *perpetuum mobile*, it is at rest."[118] The basic idea of deriving all of mechanics from this law, said Hilbert, was first introduced by Simon Stevin, in his law of equilibrium for objects in a slanted plane, but it was not clear to Stevin that what was actually involved was the reduction of the law to simpler axioms. The axiom was so absolutely obvious to Stevin, claimed Hilbert, that he had thought that a proof of it could be found without starting from any simpler assumptions.

From Hilbert's principle of conservation of energy, one can also derive the virtual velocities of the system, by adding a new axiom, namely, the principle of d'Alembert. This is done by placing in the equilibrium conditions, instead of the components

---

116 "Ist ein System in Ruhe und die Kräftefunction konstant (wirken keine Kräfte), so bleibt es in Ruhe." (Hilbert 1905a, 137)

117 "Es lässt sich zeigen, daß unter allen den Bedingungen, die die Gleichgewichtssätze der Mechanik liefern, wirklich Gleichgewicht eintritt." (Hilbert 1905a, 138)

118 "Es folgt jedoch nicht, daß diese Bedingungen auch *notwendig* für das Gleichgewicht sind, daß nicht etwa auch unter andern Umständen ein mechanisches System im Gleichgewicht sein kann. Es muß also noch ein Axiom hinzugenommen werden, des Inhaltes etwa: Ein mechanisches System kann *nur* dann im Gleichgewicht sein, wenn es dem Axiom von <der Unmöglichkeit des> Perpetuum mobile gemäß in Ruhe ist." (Hilbert 1905a, 138)

X,Y,Z of a given force-field acting on every mass point, the expressions $X-mx'',Y-my'',Z-mz''$.. In other words, the principle establishes that motion takes place in such a way that at every instant of time, equilibrium obtains between the force and the acceleration. In this case we obtain a very systematic and simple derivation of the Lagrangian equations, and therefore of the whole of mechanics, from three axioms: the two connected with the principle of conservation of energy (as sufficient and necessary conditions) and d'Alembert's principle, added now.

A third way to derive mechanics is based on the concept of impulse. Instead of seeing the force field K as a continuous function of t, we consider K as first null, or of a very small value; then, suddenly, as increasing considerably in a very short interval, from t to $t+\tau$, and finally decreasing again suddenly. If one considers this kind of process at the limit, namely, when $\tau=0$ one then obtains an impulse, which does not directly influence the acceleration, like a force, but rather creates a sudden velocity-change. The impulse is a time-independent vector, which however acts at a given point in time: at different points in time, different impulses may take place. The law that determines the action of an impulse is expressed by Bertrand's principle. This principle imposes certain conditions on the kinetic energy, so that it directly yields the velocity. It states that:

> The kinetic energy of a system set in motion as a consequence of an impulse must be maximal, as compared to the energies produced by all motions admissible under the principle of conservation of energy.[119]

The law of conservation is invoked here in order to establish that the total energy of the system is the same before and after the action of the impulse.

Bertrand's principle, like the others, could also be deduced from the elaborated body of mechanics by applying a limiting process. To illustrate this idea, Hilbert resorted to an analogy with optics: the impulse corresponds to the discontinuous change of the refraction coefficients affecting the velocity of light when it passes through the surface of contact between two media. But, again, as with the other alternative principles of mechanics, we could also begin with the concept of impulse as the basic one, in order to derive the whole of mechanics from it. This alternative assumes the possibility of constructing mechanics without having to start from the concept of force. Such a construction is based on considering a sequence of successive small impulses in arbitrarily small time-intervals, and in recovering, by a limiting process, the continuous action of a force. This process, however, necessitates the introduction of the continuity axiom discussed above. In this way, finally, the whole of mechanics is reconstructed using only two axioms: Bertrand's principle and the said axiom of continuity. In fact, this assertion of Hilbert is somewhat misleading, since his very formulation of Bertrand's principle presupposes the acceptance of the law of conservation of energy. In any case, Hilbert believed that also in this case, a

119 "Nach einem Impuls muß die kinetische Energie des Systems bei der <wirklich> eintretenden Bewegung ein Maximum sein gegenüber allen mit dem Satze von der Erhaltung der Energie verträglichen Bewegungen." (Hilbert 1905a, 141)

completely analogous process could be found in the construction of geometric optics: first one considers the process of sudden change of optical density that takes place in the surface that separates two media; then, one goes in the opposite direction, and considers, by means of a limiting process, the passage of a light ray through a medium with continuously varying optical density, seeing it as a succession of many infinitely small, sudden changes of density.

Another standard approach to the foundations of mechanics that Hilbert discussed is the one based on the use of the Hamiltonian principle as the only axiom. Consider a force field K and a potential scalar function U such that K is the gradient of U. If T is the kinetic energy of the system, then Hamilton's principle requires that the motion of the system from a given starting point, at time t_1, and an endpoint, at time t_2, takes place along the path that makes the integral

$$\int_{t_1}^{t_2} (T - U)dt$$

an *extremum* among all possible paths between those two points. The Lagrangian equations can be derived from this principle, and the principle is valid for continuous as well as for discrete masses. The principle is also valid for the case of additional constraints, insofar as these constraints do not contain differential quotients that depend on the velocity or on the direction of motion (non-holonomic conditions). Hilbert added that Gauss's principle was valid for this exception.

Hilbert's presentation of mechanics so far focused on approaches that had specifically been criticized by Hertz: the traditional one, based on the concepts of time, space, mass and force, and the energetic one, based on the use of Hamilton's principle. To conclude this section, Hilbert proceeded to discuss the approaches to the foundations of mechanics introduced in the textbooks of Hertz and Boltzmann respectively. Hilbert claimed that both intended to simplify mechanics, but each from an opposite perspective.

Expressing once again his admiration for the perfect Euclidean structure of Hertz's construction of mechanics,[120] Hilbert explained that for Hertz, all the effects of forces were to be explained by means of rigid connections between bodies; but he added that this explanation did not make clear whether one should take into account the atomistic structure of matter or not. Hertz's only axiom, as described by Hilbert, was the principle of the straightest path (*Das Prinzip von der geradesten Bahn*), which is a special case of the Gaussian principle of minimal constraint, for the force-free case. According to Hilbert, Hertz's principle is obtained from Gauss's by substituting in the place of the parameter t, the arc lengths s of the curve. The curvature

120 "Er liefert jedenfalls von dieser Grundlage aus in abstrakter und präcisester Weise einen wunderbaren Aufbau der Mechanik, indem er ganz nach Euklidischem Ideale ein vollständiges System von Axiomen und Definitionen aufstellt." (Hilbert 1905a, 146)

$$m.\left\{\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2+\left(\frac{d^2z}{ds^2}\right)^2\right\}$$

of the path is to be minimized, in each of its points, when compared with all the other possible paths in the same direction that satisfy the constraint. On this path, the body moves uniformly if one also assumes Newton's first law.[121] In fact, this requirement had been pointed out by Hertz himself in the introduction to the Principles. As one of the advantages of his mathematical formulation, Hertz mentioned the fact that he does not need to assume, with Gauss, that nature intentionally keeps a certain quantity (the constraint) as small as possible. Hertz felt uncomfortable with such assumptions.

Boltzmann, contrary to Hertz, intended to explain the constraints and the rigid connections through the effects of forces, and in particular, of central forces between any two mass points. Boltzmann's presentation of mechanics, according to Hilbert, was less perfect and less fully elaborated than that of Hertz.

In discussing the principles of mechanics in 1905, Hilbert did not explicitly separate differential and integral principles. Nor did he comment on the fundamental differences between the two kinds. He did so, however, in the next winter semester, in a course devoted exclusively to mechanics (Hilbert 1905–6, §3.1.2).[122]

Hilbert closed his discussion on the axiomatics of mechanics with a very interesting, though rather speculative, discussion involving Newtonian astronomy and continuum mechanics, in which methodological and formal considerations led him to ponder the possibility of unifying mechanics and electrodynamics. It should be remarked that neither Einstein's nor Poincaré's 1905 articles on the electrodynamics of moving bodies is mentioned in any of Hilbert's 1905 lectures; most likely, Hilbert was not aware of these works at the time.[123] Hilbert's brief remarks here, on the other hand, strongly bring to mind the kind of argument, and even the notation, used by Minkowski in his first public lectures on these topics in 1907 in Göttingen.

Earlier presentations of mechanics, Hilbert said, considered the force—expressed in terms of a vector field—as given, and then investigated its effect on motion. In Boltzmann's and Hertz's presentations, for the first time, force and motion were con-

---

121 "Die Bewegung eines jeden Systemes erfolgt gleichförmig in einer '*geradesten Bahn*', d.h. für einen Punkt: die Krümmung

$$m.\left\{\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2+\left(\frac{d^2z}{ds^2}\right)^2\right\}$$

der Bahnkurve soll ein Minimum sein, in jedem Orte, verglichen mit allen andern den Zwangsbedingungen gehorchenden Bahnen derselben Richtung, und auf dieser Bahn bewegt sich der Punkt gleichförmig." (Hilbert 1905a, 146–147)

122 The contents of this course are analyzed in some detail in (Blum 1994).

123 This particular lecture of Hilbert is dated in the manuscript 26 July 1905, whereas Poincaré's article was submitted for publication on 23 July 1905, and Einstein's paper three weeks later. Poincaré had published a short announcement on 5 June 1905, in the *Comptes rendus* of the Paris Academy of Sciences.

sidered not as separate, but rather as closely interconnected and mutually interacting, concepts. Astronomy is the best domain in which to understand this interaction, since Newtonian gravitation is the only force acting on the system of celestial bodies. In this system, however, the force acting on a mass point depends not only on its own position but also on the positions and on the motions of the other points. Thus, the motions of the points and the acting forces can only be determined simultaneously. The potential energy in a Newtonian system composed of two points $(a|b|c)$ and $(x|y|z)$ equals, as it is well-known, $-\frac{1}{r_{\substack{a,b,c\\x,y,z}}}$, the denominator of this fraction being the distance between the two points. This is a symmetric function of the two points, and thus it conforms to Newton's law of the equality of action and reaction. Starting from these general remarks, Hilbert went on to discuss some ideas that, he said, came from an earlier work of Boltzmann and which might lead to interesting results. Which of Boltzmann's works Hilbert was referring to here is not stated in the manuscript. However, from the ensuing discussion it is evident that Hilbert had in mind a short article by Boltzmann concerning the application of Hertz's perspective to continuum mechanics (Boltzmann 1900).

Hertz himself had already anticipated the possibility of extending his point of view from particles to continua. In 1900 Richard Reiff (1855–1908) published an article that developed this direction (Reiff 1900), and soon Boltzmann published a reply pointing out an error. Boltzmann indicated, however, that Hertz's point of view could be correctly extended to include continua, the possibility seemed to arise of constructing a detailed account of the whole world of observable phenomena.[124] Boltzmann meant by this that one could conceivably follow an idea developed by Lord Kelvin, J.J. Thomson and others, that considered atoms as vortices or other similar stationary motion phenomena in incompressible fluids; this would offer a concrete representation of Hertz's concealed motions and could provide the basis for explaining all natural phenomena. Such a perspective, however, would require the addition of many new hypotheses which would be no less artificial than the hypothesis of action at a distance between atoms, and therefore—at least given the current state of physical knowledge—little would be gained by pursuing it.

Boltzmann's article also contained a more positive suggestion, related to the study of the mechanics of continua in the spirit of Hertz. Following a suggestion of Brill, Boltzmann proposed to modify the accepted Eulerian approach to this issue. The latter consisted in taking a fixed point in space and deriving the equations of motion of the fluid by studying the behavior of the latter at the given point. Instead of this Boltzmann suggested a Lagrangian approach, deducing the equations by looking at an element of the fluid as it moves through space. This approach seemed to Boltzmann to be the natural way to extend Hertz's point of view from particles to continua,

124 "... ein detailliertes Bild der gesamten Erscheinungswelt zu erhalten." (Boltzmann 1900, 668)

and he was confident that it would lead to the equations of motion of an incompressible fluid as well as to those of a rigid body submerged in such a fluid.[125] In 1903 Boltzmann repeated these ideas in a seminar taught in Vienna, and one of his students decided to take the problem as the topic of his doctoral dissertation of 1904: this was Paul Ehrenfest (1880–1993). Starting from Boltzmann's suggestion, Ehrenfest studied various aspects of the mechanics of continua using a Lagrangian approach. In fact, Ehrenfest in his dissertation used the terms Eulerian and Lagrangian with the meaning intended here, as Boltzmann in his 1900 article had not (Ehrenfest 1904, 4–5). The results obtained in the dissertation helped to clarify the relations between the differential and the integral variational principles for non-holonomic systems, but they offered no real contribution to an understanding of all physical phenomena in terms of concealed motions and masses, as Boltzmann and Ehrenfest may have hoped.[126]

Ehrenfest studied in Göttingen between 1901 and 1903, and returned there in 1906 for one year, before moving with his mathematician wife Tatyana to St. Petersburg. We don't know the details of Ehrenfest's attendance at Hilbert's lectures during his first stay in Göttingen. Hilbert taught courses on the mechanics of continua in the winter semester of 1902–1903 and in the following summer semester of 1903, which Ehrenfest may well have attended. Nor do we know whether Hilbert knew anything about Ehrenfest's dissertation when he taught his course in 1905. But be that as it may, at this point in his lectures, Hilbert connected his consideration of Newtonian astronomy to the equations of continuum mechanics, while referring to the dichotomy between the Lagrangian and the Eulerian approach, and using precisely those terms. Interestingly enough, the idea that Hilbert pursued in response to Boltzmann's article was not that the Lagrangian approach would be the natural one for studying mechanics of continua, but rather the opposite, namely, that a study of the continua following the Eulerian approach, and assuming an atomistic worldview, could lead to a unified explanation of all natural phenomena.

Consider a free system subject only to central forces acting between its mass-points —and in particular only forces that satisfy Newton's law, as described above. An axiomatic description of this system would include the usual axioms of mechanics, together with the Newtonian law as an additional one. We want to express this system, said Hilbert, as concisely as possible by means of differential equations. In the most general case we assume the existence of a continuous mass distribution in space, $\rho = (x, y, z, t)$. In special cases we have $\rho = 0$ within a well-delimited region; the case of astronomy, in which the planets are considered mass-points, can be derived from this special case by a process of passage to the limit. Hilbert explained what the Lagrangian approach to this problem would entail. That approach, he added, is the most appropriate one for discrete systems, but often it is also conveniently used in the mechanics of continua. Here, however, he would follow the Eulerian approach

125 For more details, cf. (Klein 1970, 64–66).

126 For details on Ehrenfest's dissertation, see (Klein 1970, 66–74).

to derive equations of the motion of a unit mass-particle in a continuum. The ideas discussed in this section, as well as in many other parts of this course, hark back to those he developed in somewhat greater technical detail in his 1902–1903 course on continuum mechanics, but here a greater conceptual clarity and a better understanding of the possible, underlying connections across disciplines is attained, thanks to the systematic use of an axiomatic approach in the discussion.

Let V denote the velocity of the particle at time t and at coordinates (x, y, z) in the continuum. V has three components $u = u(x, y, z, t)$, v and w. The acceleration vector for the unit particle is given by dV/dt, which Hilbert wrote as follows:[127]

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + u\frac{\partial V}{\partial x} + v\frac{\partial V}{\partial y} + w\frac{\partial V}{\partial z} = \frac{\partial V}{\partial t} + V \times \operatorname{curl} V - \frac{1}{2}\operatorname{grad}(V.V).$$

Since the only force acting on the system is Newtonian attraction, the potential energy at a point $(x|y|z)$ is given by

$$P = -\iiint -\frac{\rho'}{r_{\substack{x',y',z' \\ x,y,z}}} dx'dy'dz'$$

where ρ' is the mass density at the point $(x'|y'|z')$. The gradient of this potential equals the force acting on the particle, and therefore we obtain three equations of motion that can succinctly be expressed as follows:

$$\frac{\partial V}{\partial t} + V \times \operatorname{curl} V - \frac{1}{2}\operatorname{grad}(V.V) = \operatorname{grad} P.$$

One can add two additional equations to these three. First, the Poisson equation, which Hilbert calls "potential equation of Laplace":

$$\Delta P = 4\pi\rho$$

where Δ denotes the Laplacian operator (currently written as ∇^2). Second, the constancy of the mass in the system is established by means of the continuity equation:[128]

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}(\rho \cdot V).$$

We have thus obtained five differential equations involving five functions (the components u, v, w of V, P and ρ of the four variables x, y, z, t. The equations are

127 In the manuscript the formula in the leftmost side of the equation appears twice, having a "-" sign in front of $V \times \operatorname{curl} V$. This is obviously a misprint, as a straightforward calculation readily shows.

128 In his article mentioned above, Reiff had tried to derive the pressure forces in a fluid starting only from the conservation of mass (Reiff 1900). Boltzmann pointed out that Reiff had obtained a correct result because of a compensation error in his mathematics. See (Klein 1970, 65).

completely determined when we know their initial values and other boundary conditions, such as the values of the functions at infinity. Hilbert called the five equations so obtained the "Newtonian world-functions," since they account in the most general way and in an axiomatic fashion for the motion of the system in question: a system that satisfies the laws of mechanics and the Newtonian gravitational law. It is interesting that Hilbert used the term "world-function" in this context, since the similar ones "world-point" and "world-postulate," were introduced in 1908 by Minkowski in the context of his work on electrodynamics and the postulate of relativity. Unlike most of the mathematical tools and terms introduced by Minkowski, this particular aspect of his work was not favorably received, and is hardly found in later sources (with the exception of "world-line"). Hilbert, however, used the term "world-function" not only in his 1905 lectures, but also again in his 1915 work on general relativity, where he again referred to the Lagrangian function used in the variational derivation of the gravitational field equations as a "world-function."

Besides the more purely physical background to the issues raised here, it is easy to detect that Hilbert was excited about the advantages and the insights afforded by the vectorial formulation of the Eulerian equations. Vectorial analysis as a systematic way of dealing with physical phenomena was a fairly recent development that had crystallized towards the turn of the century, mainly through its application by Heaviside in the context of electromagnetism and through the more mathematical discussion of the alternative systems by Gibbs.[129] The possibility of extending its use to disciplines like hydrodynamics had arisen even more recently, especially in the context of the German-speaking world. Thus, for instance, the *Encyklopädie* article on hydrodynamics, written in 1901, still used the pre-vectorial notation (Love 1901, 62–63).[130] Only one year before Hilbert's course, speaking at the International Congress of Mathematicians in Heidelberg, the Göttingen applied mathematician Ludwig Prandtl still had to explain to his audience how to write the basic equations of hydrodynamics "following Gibbs's notation" (Prandtl 1904, 489). Among German textbooks on vectorial analysis of the turn of the century,[131] formulations of the Eulerian equations like that quoted above appear in Alfred Heinrich Bucherer's textbook of 1903 (Bucherer 1903, 77–84) and in Richard Gans's book of 1905 (Gans 1905, 66–67). Whether he learnt about the usefulness of the vectorial notation in this context from his colleague Prandtl or from one of these textbooks, Hilbert was certainly impressed by the unified perspective it afforded from the formal point of view. Moreover, he seems also to have wanted to deduce far-reaching physical conclusions from this formal similarity. Hilbert pointed out in his lectures the strong analogy between this formulation of the equations and Maxwell's equations of electrodynamics, though in the latter we have two vectors E, and B, the electric and the magnetic fields, against only one here, V.

129 Cf. (Crowe 1967, 182–224).

130 The same is the case for (Lamb 1895, 7). This classical textbook, however, saw many later editions in which the vectorial formulation was indeed adopted.

131 Cf. (Crowe 1967, 226–233).

He also raised the following question: can one obtain the whole of mechanics starting from these five partial equations as a single axiom, or, if that is not the case, how far can its derivation in fact be carried? In other words: if we want to derive the whole of mechanics, to what extent can we limit ourselves to assuming only Newtonian attraction or the corresponding field equations?[132] It would also be interesting, he said, to address the question of how far the analogy of gravitation with electrodynamics can be extended. Perhaps, he said, one can expect to find a formula that simultaneously encompasses these five equations and the Maxwellian ones together. This discussion of a possible unification of mechanics and electrodynamics also echoed, of course, the current foundational discussion that I have described in the preceding sections. It also anticipates what will turn out to be one of the pillars of Hilbert's involvement with general relativity in 1915.

Hilbert's reference to Hertz and Boltzmann in this context, and his silence concerning recent works of Lorentz, Wien, and others, is the only hint he gave in his 1905 lectures as to his own position on the foundational questions of physics. In fact, throughout these lectures Hilbert showed little inclination to take a stand on physical issues of this kind. Thus, his suggestion of unifying the equations of gravitation and electrodynamics was advanced here mainly on methodological grounds, rather than expressing, at this stage at least, any specific commitment to an underlying unified vision of nature. At the same time, however, his suggestion is quite characteristic of the kind of mathematical reasoning that would allow him in later years to entertain the possibility of unification and to develop the mathematical and physical consequences that could be derived from it.

6.2 Kinetic Theory of Gases

A main application of the calculus of probabilities that Hilbert considered is in the kinetic theory of gases. He opened this section by expressing his admiration for the remarkable way this theory combined the postulation of far-reaching assumptions about the structure of matter with the use of probability calculus, a combination that had been applied in a very illuminating way, leading to new physical results. Several works that appeared by end of the nineteenth century had changed the whole field of the study of gases, thus leading to a more widespread appreciation of the value of the statistical approach. The work of Planck, Gibbs and Einstein attracted a greater interest in and contributed to an understanding of Boltzmann's statistical interpretation of entropy.[133]

---

132 "Es wäre nun die Frage, ob man mit ~~einem~~ diesen 5 partiellen Gleichungen als einzigem Axiom nicht auch überhaupt in der Mechanik auskommt, oder wie weit das geht, d.h. wie weit man sich auf Newtonsche Attraktion bezw. auf die entsprechenden Feldgleichungen beschränken kann." (Hilbert 1905a, 154)

133 Kuhn (1978, 21) quotes in this respect the well-known textbook, (Gibbs 1902), and an "almost forgotten" work, (Einstein 1902).

It is easy to see, then, why Hilbert would have wished to undertake an axiomatic treatment of the kinetic theory of gases: not only because it combined physical hypotheses with probabilistic reasoning in a scientifically fruitful way, as Hilbert said in these lectures, but also because the kinetic theory was a good example of a physical theory where, historically speaking, additional assumptions had been gradually added to existing knowledge without properly checking the possible logical difficulties that would arise from this addition. The question of the role of probability arguments in physics was not settled in this context. In Hilbert's view, the axiomatic treatment was the proper way to restore order to this whole system of knowledge, so crucial to the contemporary conception of physical science.

In stating the aim of the theory as the description of the macroscopic states of a gas, based on statistical considerations about the molecules that compose it, Hilbert assumed without any further comment the atomistic conception of matter. From this picture, he said, one obtains, for instance, the pressure of the gas as the number of impacts of the gas molecules against the walls of its container, and the temperature as the square of the sum of the mean velocities. In the same way, entropy becomes a magnitude with a more concrete physical meaning than is the case outside the theory. Using Maxwell's velocity distribution function, Boltzmann's logarithmic definition of entropy, and the calculus of probabilities, one obtains the law of constant increase in entropy. Hilbert immediately pointed out the difficulty of combining this latter result with the reversibility of the laws of mechanics. He characterized this difficulty as a paradox, or at least as a result not yet completely well established.[134] In fact, he stressed that the theory had not yet provided a solid justification for its assumptions, and ever new ideas and stimuli were constantly still being added.

Even if we knew the exact position and velocities of the particles of a gas— Hilbert explained—it is impossible in practice to integrate all the differential equations describing the motions of these particles and their interactions. We know nothing of the motion of individual particles, but rather consider only the average magnitudes that are dealt with by the probabilistic kinetic theory of gases. In an oblique reference to Boltzmann's replies, Hilbert stated that the combined use of probabilities and infinitesimal calculus in this context is a very original mathematical contribution, which may lead to deep and interesting consequences, but which at this stage has in no sense been fully justified. Take, for instance, one of the well-known results of the theory, namely, the equations of vis viva. In the probabilistic version of the theory, Hilbert said, the solution of the corresponding differential equation does not emerge solely from the differential calculus, and yet it is correctly determined. It might conceivably be the case, however, that the probability calculus could have contradicted well-known results of the theory, in which case, using that calculus would clearly yield what would be considered unacceptable conclusions. Hilbert explained this

134 "Hier können wir aber bereits ein paradoxes, zum mindesten nicht recht befriedigendes Resultat feststellen." (Hilbert 1905a, 176)

warning by showing how a fallacious probabilistic argument could lead to contradiction in the theory of numbers.

Take the five classes of congruence module 5 in the natural numbers, and consider how the prime numbers are distributed among these classes. For any integer x, let $A(x)$ be the number of prime numbers which are less than x, and let $A_0(x),\ldots,A_4(x)$, be the corresponding values of the same function, when only the numbers in each of the five classes are considered. Using the calculus of probabilities in a similar way to that used in the integration of the equations of motion of gas particles, one could reason as follows: The distribution of prime numbers is very irregular, but according to the laws of probability, this irregularity is compensated if we just take a large enough quantity of events. In particular, the limits at infinity of the quotients $A_i(x)/A(x)$ are all equal for $i=0,\ldots,4$, and therefore equal to $1/5$. But it is clear, on the other hand, that in the class of numbers of the form $5m$, there are no prime numbers, and therefore $A_0(x)/A(x)=0$. One could perhaps correct the argument by limiting its validity to the other four classes, and thus conclude that:

$$\lim_{x=\infty} \frac{A_i(x)}{A(x)} = \frac{1}{4}, \text{ for } i=1,2,3,4.$$

Although this latter result is actually correct, Hilbert said, one cannot speak here of a real proof. The latter could only be obtained through deep research in the theory of numbers. Had we not used here the obvious number-theoretical fact that $5m$ can never be a prime number, we might have been misled by the probabilistic proof. Something similar happens in the kinetic theory of gases, concerning the integration of the vis viva. One assumes that Maxwell's distribution of velocities obeys a certain differential equation of mechanics, and in this way a contradiction with the known value of the integral of the vis viva is avoided. Moreover, according to the theory, because additional properties of the motion of the gas particles, which are prescribed by the differential equations, lie very deep and are only subtly distinguishable, they do not affect relatively larger values, such as the averages used in the Maxwell laws.[135] As in the case of the prime numbers, however, Hilbert did not consider this kind of reasoning to be a real proof.

All this discussion, which Hilbert elaborated in further detail, led him to formulate his view concerning the role of probabilistic arguments in mathematical and physical theories. In this view, surprisingly empiricist and straightforwardly formu-

135 "Genau so ist es nun hier in der kinetischen Gastheorie. Indem wir behaupten, daß die Maxwellsche Geschwindigkeitsverteilung den mechanischen Differentialgleichnungen genügt, vermeiden wir wohl einen Verstoß gegen das sofort bekannte Integral der lebendigen Kraft; weiterhin aber wird die Annahme gemacht, ~~daß~~ die durch die Differentialgleichungen geforderten weiteren Eigenschaften der Gaspartikelbewegung liegen soviel tiefer und sind so feine Unterscheidungen, daß sie so grobe Aussagen über mittlere Werte, wie die des Maxwellschen Gesetzes, nicht berühren." (Hilbert 1905a, 180–181)

lated, the calculus of probability is not an exact mathematical theory, but one that may appropriately be used as a first approximation, provided we are dealing with immediately apparent mathematical facts. Otherwise it may lead to significant contradictions. The use of the calculus of probabilities is justified—Hilbert concluded—insofar as it leads to results that are correct and in accordance with the facts of experience or with the accepted mathematical theories.[136]

Beginning in 1910 Hilbert taught courses on the kinetic theory of gases and on related issues, and also published original contributions to this domain. In particular, as part of his research on the theory of integral equations, which began around 1902, he solved in 1912 the so-called Boltzmann equation.[137]

6.3 Electrodynamics

The manuscript of the lecturer indicates that Hilbert did not discuss electrodynamics before 14 July 1905. By that time Hilbert must have been deeply involved with the issues studied in the electron-theory seminar. These issues must surely have appeared in the lectures as well, although the rather elementary level of discussion in the lectures differed enormously from the very advanced mathematical sophistication characteristic of the seminar. As mentioned above, at the end of his lectures on mechanics Hilbert had addressed the question of a possible unification of the equations of gravitation and electrodynamics, mainly based on methodological considerations. Now he stressed once more the similarities underlying the treatment of different physical domains. In order to provide an axiomatic treatment of electrodynamics similar to those of the domains discussed above—Hilbert opened this part of his lectures—one needs to account for the motion of an electron by describing it as a small electrified sphere and by applying a process of passage to the limit.

One starts therefore by considering a material point m in the classical presentation of mechanics. The kinetic energy of a mass-point is expressed as

$$L(v) = \frac{1}{2}mv^2.$$

The derivatives of this expression with respect to the components v_s of the velocity v define the respective components of the momentum

$$\frac{\partial L(v)}{\partial v_s} = m.v_s.$$

---

136 "… sie ist keine exakte mathematische Theorie, aber zu einer ersten Orientierung, wenn man nur alle unmittelbar leicht ersichtlichen mathematischen Tatsachen benutzt, häufig sehr geeignet; sonst führt sie sofort zu groben Verstößen. Am besten kann man wohl immer nachträglich sagen, daß die Anwendung der Wahrscheinlichkeitsrechnung immer dann berechtigt und erlaubt ist, wo sie zu richtigen, mit der Erfahrung bezw. der sonstigen mathematischen Theorie übereinstimmenden Resultaten führt." (Hilbert 1905a, 182–183)

137 In (Hilbert 1912a, chap. XXII).

If one equates the derivative of the latter with respect to time to the components of the forces—seen as the negative of the partial derivatives of the potential energy—one gets the equations of motion:

$$d\frac{\partial}{\partial v_s}\Big/dt + \frac{\partial U}{\partial s} = 0 \qquad s = (x, y, z).$$

As was seen earlier in the lectures on mechanics, an alternative way to attain these equations is to use the functions L, U and the variational equation characteristic of the Hamiltonian principle:

$$\int_{t_1}^{t_2} (L - U)\,dt = \text{Minim.}$$

This principle can be applied, as Laplace did in his *Celestial Mechanics*, even without knowing anything about L, except that it is a function of the velocity. In order to determine the actual form of L, one must then introduce additional axioms. Hilbert explained that in the context of classical mechanics, Laplace had done this simply by asserting what for him was an obvious, intuitive notion concerning relative motion, namely, that we are not able to perceive any uniform motion of the whole universe.[138] From this assumption Laplace was able to derive the actual value $L(v) = (1/2)mv^2$. This was for Hilbert a classical instance of the main task of the axiomatization of a physical science, as he himself had been doing throughout his lectures for the cases of the addition of vectors, thermodynamics, insurance mathematics, etc.: namely, to formulate the specific axiom or axioms underlying a particular physical theory, from which the specific form of its central, defining function may be derived. In this case, Laplace's axiom is nothing but the expression of the Galilean invariance of the Newtonian laws of motion although Hilbert did not use this terminology here.

In the case of the electron, as Hilbert had perhaps recently learnt in the electron-theory seminar, this axiom of Galilean invariance, is no longer valid, nor is the specific form of the Lagrangian function. Yet—and this is what Hilbert stressed as a remarkable fact—the equation of motion of the electron can nevertheless be derived following considerations similar to those applied in Laplace's case. One need only find the appropriate axiom to effect the derivation. Without further explanation, Hil-

138 "Zur Festlegung von L muß man nun natürlich noch Axiome hinzunehmen, und Laplace kommt da mit einer allgemeinen, ihm unmittelbar anschaulichen Vorstellung über Relativbewegung aus, daß wir nämlich eine gleichförmige Bewegung des ganzen Weltalls nicht merken würden. Alsdann läßt sich die Form $mv^2/2$ von $L(v)$ bestimmen, und das ist wieder die ganz analoge Aufgabe zu denen, die das Fundament der Vektoraddition, der Thermodynamik, der Lebensversicherungsmathematik u.a. bildeten." (Hilbert 1905a, 187)

bert wrote down the Lagrangian that describes the motion of the electron. This may be expressed as

$$L(v)= \mu\frac{1-v^2}{v}\cdot\log\frac{1+v}{1-v}$$

where v denotes the ratio between the velocity of the electron and the speed of light, and μ is a constant, characteristic of the electron and dependent on its charge. This Lagrangian appears, for instance, in Abraham's first article on the dynamics of the electron, and a similar one appears in the article on Lorentz's *Encyklopädie* article.[139] If not earlier than that, Hilbert had studied these articles in detail in the seminar, where Lorentz's article was used as a main text.

If, as in the case of classical mechanics, one again chooses to consider the differential equation or the corresponding variational equation as the single, central axiom of electron theory, taking L as an undetermined function of v whose exact expression one seeks to derive, then—Hilbert said—in order to do so, one must introduce a specific axiom, characteristic of the theory and as simple and plausible as possible. Clearly—he said concluding this section—this theory will require more, or more complicated, axioms than the one introduced by Laplace in the case of classical mechanics.[140] The electron-theory seminar had been discussing many recent contributions, by people such as Poincaré, Lorentz, Abraham and Schwarzschild, who held conflicting views on many important issues. It was thus clear to Hilbert that, at that point in time at least, it would be too early to advance any definite opinion as to the specific axiom or axioms that should be placed at the basis of the theory. This fact, however, should not affect in principle his argument as to how the axiomatic approach should be applied to the theory.

It is noteworthy that in 1905 Hilbert did not mention the Lorentz transformations, which were to receive very much attention in his later lectures on physics. Lorentz published the transformations in an article of 1904 (Lorentz 1904b), but this article was not listed in the bibliography of the electron theory seminar,[141] and it is likely that Hilbert was not aware of it by the time of his lectures.

6.4 A post-1909 addendum

To conclude this account of the 1905 lectures, it is interesting to notice that several years after having taught the course, Hilbert returned to the manuscript and added

---

139 Respectively, (Abraham 1902, 37; Lorentz 1904a, 184). Lorentz's Lagrangian is somewhat different, since it contains two additional terms, involving the inverse of v^3.

140 "Nimmt man nun wieder die Differentialgleichungen bzw. das zugehörige Variationsproblem als Axiom und läßt L zunächst als noch unbestimmte Funktion von v stehen, so handelt es sich darum, dafür möglichst einfache und plausible Axiome so zu konstruiren, daß sie gerade jene Form von $L(v)$ bestimmen. Natürlich werden wir mehr oder kompliciertere Axiome brauchen, als in dem einfachen Falle der Mechanik bei Laplace." (Hilbert 1905a, 188)

141 Cf. (Pyenson 1979, 103).

some remarks on the front page in his own handwriting. He mentioned two more recent works he thought relevant to understanding the use of the axiomatic method in physics. First, he referred to a new article by Hamel on the principles of mechanics. Hamel's article, published in 1909, contained philosophical and critical remarks concerning the issues discussed in his own earlier article of 1905 (the one mentioned by Hilbert with reference to the axiomatization of vector addition). In particular, it discussed the concepts of absolute space, absolute time and force, as a priori concepts of mechanics. The contents of this article are beyond the scope of our discussion here. Hilbert's interest in it may have stemmed from a brief passage where Hamel discussed the significance of Hilbert's axiomatic method (Hamel 1909, 358). More importantly perhaps, it also contained an account of a new system of axioms for mechanics.[142]

Second, in a formulation that condenses in a very few sentences his understanding of the principles and goals of axiomatization, as they apply to geometry and to various domains of physics, Hilbert also directed attention to what he saw as Planck's application of the axiomatic method in the latter's recent research on quantum theory. Hilbert thus wrote:

> It is of special interest to notice how the axiomatic method is put to use by Planck—in a more or less consistent and in a more or less conscious manner—even in modern quantum theory, where the basic concepts have been so scantily clarified. In doing this, he sets aside electrodynamics in order to avoid contradiction, much as, in geometry, continuity is set aside in order to remove the contradiction in non-Pascalian geometry, or like, in the theory of gases, mechanics is set aside in favor of the axiom of probability (maximal entropy), thus applying only the Stossformel or the Liouville theorem, in order to avoid the objections involved in the reversibility and recurrence paradoxes.[143]

From this remark we learn not only that Hilbert was aware of the latest advances in quantum theory (though, most probably, not in great detail) but also that he had a good knowledge of recent writings of Paul and Tatyana Ehrenfest. Beginning in 1906 the Ehrenfests had made important contributions to clarifying Boltzmann's ideas in a series of publications on the conceptual foundations of statistical mechanics. The two last terms used by Hilbert in his hand-written remark (*Umkehr- oder Wiederkehreinwand)* were introduced only in 1907 by them, and were made widely known only

142 According to Clifford Truesdell (1968, 336), this article of Hamel, together with the much later (Noll 1959), are the "only two significant attempts to solve the part of Hilbert's sixth problem that concern mechanics [that] have been published." One should add to this list at least another long article (Hamel 1927) that appeared in vol. 5 of the *Handbuch der Physik*.

143 Hilbert (1905a), added "<Besonders interessant ist es zu sehen, wie die axiomatische Methode von Planck sogar bei der modernen Quantentheorie, wo die Grundbegriffe noch so wenig geklärt sind, in mehr oder weniger konsequenter und in mehr oder weniger bewusster Weise zur Anwendung gebracht werden: dabei Ausschaltung der Elektrodynamik, um Widerspruch zu vermeiden—gerade wie in der Geometrie Ausschaltung der Stetigkeit, um den Widerspruch gegen die Nichtpaskalsche Geometrie zu beseitigen, oder in der Gastheorie Ausschaltung der Mechanik (Benutzung allein der Stossformel oder des Liouvilleschen Satzes) dafür Axiom der Wahrscheinlichkeit—(Entropie Maximum), um den Widerspruch gegen den Umkehr- oder Wiederkehreinwand zu beseitigen.>"

through their *Encyklopädie* article that appeared in 1912. Hilbert may have known the term earlier from their personal contact with them, or through some other colleague.[144] Also, the *Stossformel* that Hilbert mentioned here referred probably to the *Stossanzahlansatz*, whose specific role in the kinetic theory, together with that of the Liouville theorem (that is the physicists' Liouville theorem), the Ehrenfests' article definitely contributed to clarify.[145] Moreover, the clarification of the conceptual interrelation between Planck's quantum theory and electrodynamics—alluded to by Hilbert in his added remark—was also one of Paul Ehrenfest's central contributions to contemporary physics.[146]

7. THE AXIOMATIZATION PROGRAM BY 1905 – PARTIAL SUMMARY

Hilbert's 1905 cycle of lectures on the axiomatization of physics represents the culmination of a very central thread in Hilbert's early scientific career. This thread comprises a highly visible part of his published work, namely that associated with *Grundlagen der Geometrie*, but also additional elements that, though perhaps much less evident, were nevertheless prominent within his general view of mathematics, as we have seen. Hilbert's call in 1900 for the axiomatization of physical theories was a natural outgrowth of the background from which his axiomatic approach to geometry first developed. Although in elaborating the point of view put forward in the *Grundlagen der Geometrie* Hilbert was mainly driven by the need to solve certain, open foundational questions of geometry, his attention was also attracted in this context by recent debates on the role of axioms, or first principles in physics. Hertz's textbook on mechanics provided an elaborate example of a physical theory presented in strict axiomatic terms, and—perhaps more important for Hilbert—it also discussed in detail the kind of requirements that a satisfactory system of axioms for a physical theory must fulfill. Carl Neumann's analysis of the "Galilean principle of inertia"—echoes of which we find in Hilbert's own treatment of mechanics— provided a further example of the kind of conceptual clarity that one could expect to gain from this kind of treatment. The writings of Hilbert's senior colleague at Königsberg, Paul Volkmann, show that towards the end of the century questions of this kind were also discussed in the circles he moved in. Also the works of both Boltzmann and Voss provided Hilbert with important sources of information and inspiration. From his earliest attempts to treat geometry in an axiomatic fashion in order to solve the foundational questions he wanted to address in this field, Hilbert already had in mind the axiomatization of other physical disciplines as a task that could and should be pursued in similar terms.

---

144 Hilbert was most likely present when, on 13 November 1906, Paul Ehrenfest gave a lecture at the *Göttinger Mathematische Gesellschaft* on Boltzmann's H-theorem and some of the objections (*Einwände*) commonly raised against it. This lecture is reported in *Jahresbericht der Deutschen Mathematiker-Vereinigung*, Vol. 15 (1906), 593.

145 Cf. (Klein 1970, 119–140).

146 Cf. (Klein 1970, 230–257).

Between 1900 and 1905 Hilbert had the opportunity to learn much new physics. The lecture notes of his course provide the earliest encompassing evidence of Hilbert's own picture of physical science in general and, in particular, of how he thought the axiomatic analysis of individual theories should be carried out. Hilbert's physical interests now covered a broad range of issues, and he seems to have been well aware of the main open questions being investigated in most of the domains addressed. His unusual mathematical abilities allowed him to gain a quick grasp of existing knowledge, and at the same time to consider the various disciplines from his own idiosyncratic perspective, suggesting new interpretations and improved mathematical treatments. However, one must exercise great care when interpreting the contents of these notes. It is difficult to determine with exactitude the extent to which he had studied thoroughly and comprehensively all the existing literature on a topic he was pursuing. The relatively long bibliographical lists that we find in the introductions to many of his early courses do not necessarily mean that he studied all the works mentioned there. Even from his repeated, enthusiastic reference to Hertz's textbook we cannot safely infer to what extent he had read that book thoroughly. Very often throughout his career he was content when some colleague or student communicated to him the main ideas of a recent book or a new piece of research. In fact, the official assignment of many of his assistants—especially in the years to come—was precisely that: to keep him abreast of recent advances by studying in detail the research literature of a specific field. Hilbert would then, if he were actually interested, study the topic more thoroughly and develop his own ideas.

It is also important to qualify properly the extent to which Hilbert carried out a full axiomatic analysis of the physical theories he discussed. As we saw in the preceding sections, there is a considerable difference between what he did for geometry and what he did for other physical theories. In these lectures, Hilbert never actually proved the independence, consistency or completeness of the axiomatic systems he introduced. In certain cases, like vector addition, he quoted works in which such proofs could be found (significantly, works of his students or collaborators). In other cases there were no such works to mention, and—as in the case of thermodynamics—Hilbert simply *stated* that his axioms are indeed independent. In still other cases, he barely mentioned anything about independence or other properties of his axioms. Also, his derivations of the basic laws of the various disciplines from the axioms are rather sketchy, when they appear at all. Often, Hilbert simply declared that such a derivation was possible. What is clear is that Hilbert considered that an axiomatization along the lines he suggested was plausible and could eventually be fully performed following the standards established in *Grundlagen der Geometrie*.

Yet for all these qualifications, the lecture notes of 1905 present an intriguing picture of Hilbert's knowledge of physics, notable both for its breadth and its incisiveness. They afford a glimpse into a much less known side of his Göttingen teaching activity, which must certainly be taken into account in trying to understand the atmosphere that dominated this world center of science, as well as its widespread influence. More specifically, these notes illustrate in detail how Hilbert envisaged that

axiomatic analysis of physical theories could not only contribute to conceptual clarification but also prepare the way for the improvement of theories, in the eventuality of future experimental evidence that conflicted with current predictions. If one knew in detail the logical structure of a given theory and the specific role of each of its basic assumptions, one could clear away possible contradictions and superfluous additional premises that may have accumulated in the building of the theory. At the same time, one would be prepared to implement, in an efficient and scientifically appropriate way, the local changes necessary to readapt the theory to meet the implications of newly discovered empirical data, in the eventuality of such discoveries. Indeed, Hilbert's own future research in physics, and in particular his incursion into general relativity, will be increasingly guided by this conception.

The nature and use of axioms in physical theories was discussed by many of Hilbert's contemporaries, as we have seen. Each had his own way of classifying the various kinds of axioms that are actually used or should be used. Hilbert himself did not discuss any possible such classification in detail but in his lectures we do find three different kinds of axioms actually implemented. This *de facto* classification is reminiscent, above all, of the one previously found in the writings of Volkmann. In the first place, every theory is assumed to be governed by specific axioms that characterize it and only it. These axioms usually express mathematical properties establishing relations among the basic magnitudes involved in the theory. Secondly, there are certain general mathematical principles that Hilbert saw as being valid for all physical theories. In the lectures he stressed above all the "continuity axiom," providing both a general formulation and more specific ones for each theory. As an additional general principle of this kind he suggested the assumption that all functions appearing in the natural sciences should have at least one continuous derivative. Furthermore, the universal validity of variational principles as the key to deriving the main equations of physics was a central underlying assumption of all of Hilbert's work on physics, and that kind of reasoning appears throughout these lectures as well. In each of the theories he considered in his 1905 lectures, Hilbert attempted to show how the exact analytic expression of a particular function that condenses the contents of the theory in question could be effectively derived from the specific axioms of the theory, together with more general principles. On some occasions he elaborated this idea more thoroughly, while on others he simply declared that such a derivation should be possible.

There is yet a third type of axiom for physical theories that Hilbert, however, avoided addressing in his 1905 lectures. That type comprises claims about the ultimate nature of physical phenomena, an issue that was particularly controversial during the years preceding these lectures. Although Hilbert's sympathy for the mechanical worldview is apparent throughout the manuscript of the lectures, his axiomatic analyses of physical theories contain no direct reference to it. The logical structure of the theories is thus intended to be fully understood independently of any particular position in this debate. Hilbert himself would later adopt a different stance. His work on general relativity will be based directly on his adoption of the electromagnetic worldview and, beginning in 1913, a quite specific version of it, namely,

Gustav Mie's electromagnetic theory of matter. On the other hand, Hermann Minkowski's work on electrodynamics, with its seminal reinterpretation of Einstein's special theory of relativity in terms of spacetime geometry, should be understood as an instance of the kind of axiomatic analysis that Hilbert advanced in his 1905 lectures in which, at the same time, the debate between the mechanical and the electromagnetic worldviews is avoided.

When reading the manuscript of these lectures, one cannot help speculating about the reaction of the students who attended them. This was, after all, a regular course offered in Göttingen, rather than an advanced seminar. Before the astonished students stood the great Hilbert, rapidly surveying so many different physical theories, together with arithmetic, geometry and even logic, all in the framework of a single course. Hilbert moved from one theory to the other, and from one discipline to the next, without providing motivations or explaining the historical background to the specific topics addressed, without giving explicit references to the sources, without stopping to work out any particular idea, without proving any assertion in detail, but claiming all the while to possess a unified view of all these matters. The impression must have been thrilling, but perhaps the understanding he imparted to the students did not run very deep. Hermann Weyl's account of his experience as a young student attending Hilbert's course upon his arrival in Göttingen offers direct evidence to support this impression. Thus, in his obituary of Hilbert, Weyl wrote:

> In the fullness of my innocence and ignorance I made bold to take the course Hilbert had announced for that term, on the notion of number and the quadrature of the circle. Most of it went straight over my head. But the doors of a new world swung open for me, and I had not sat long at Hilbert's feet before the resolution formed itself in my young heart that I must by all means read and study what this man had written. (Weyl 1944, 614)

But the influence of the ideas discussed in Hilbert's course went certainly beyond the kind of general inspiration described here so vividly by Weyl; they had an actual influence on later contributions to physics. Besides the works of Born and Carathéodory on thermodynamics, and of Minkowski on electrodynamics, there were many dissertations written under Hilbert, as well as the articles written under the influence of his lectures and seminars. Ehrenfest's style of conceptual clarification of existing theories, especially as manifest in the famous *Encyklopädi*e on statistical mechanics, also bears the imprint of Hilbert's approach. Still, one can safely say that little work on physical theories was actually published along the specific lines of axiomatic analysis suggested by Hilbert in *Grundlagen der Geometrie*. It seems, in fact, that such techniques were never fully applied by Hilbert or by his students and collaborators to yield detailed analyses of axiomatic systems defining physical theories. Thus, for instance, in 1927 Georg Hamel wrote a long article on the axiomatization of mechanics for the *Handbuch der Physik* (Hamel 1927). Hamel did mention Hilbert's work on geometry as the model on which any modern axiomatic analysis should be based. However, his own detailed account of the axioms needed for defining mechanics as known at that time was not followed by an analysis of the independence of the axioms, based on the construction of partial models, such as Hilbert had carried out

for geometry. Similarly, the question of consistency was discussed only summarily. Nevertheless, as Hamel said, his analysis allowed for a clearer comprehension of the logical structure of all the assumptions and their interdependence.

If the 1905 lectures represent the culmination of a thread in Hilbert's early career, they likewise constitute the beginning of the next stage of his association with physics. In the next years, Hilbert himself became increasingly involved in actual research in mathematical physics and he taught many courses on various topics thus far not included within his scientific horizons.

8. LECTURES ON MECHANICS AND CONTINUUM MECHANICS

In his early courses on mechanics or continuum mechanics, Hilbert's support for the atomistic hypothesis, as the possible basis for a reductionistic, mechanical foundation of the whole of physics, was often qualified by referring to the fact that the actual attempts to provide a detailed account of how such a reduction would work in specific cases for the various physical disciplines had not been fully and successfully realized by then. Thus for instance, in his 1906 course on continuum mechanics, Hilbert described the theory of elasticity as a discipline whose subject-matter is the deformation produced on solid bodies by interaction and displacement of molecules. On first sight this would seem to be a classical case in which one might expect a direct explanation based on atomistic considerations. Nevertheless Hilbert suggested that, for lack of detailed knowledge, a different approach should be followed in this case:

> We will have to give up going here into a detailed description of these molecular processes. Rather, we will only look for those parameters on which the measurable deformation state of the body depends at each location. The form of the dependence of the Lagrangian function on these parameters will then be determined, which is actually composed by the kinetic and potential energy of the individual molecules. Similarly, in thermodynamics we will not go into the vibrations of the molecules, but we will rather introduce temperature itself as a general parameter and we will investigate the dependence of energy on it.[147]

The task of deducing the exact form of the Lagrangian under specific requirements postulated as part of the theory was the approach followed in the many examples already discussed above. This tension between reductionistic and phenomenological explanations in physics is found in Hilbert's physical ideas throughout the years and it eventually led to his abandonment of mechanical reduc-

147 "Wir werden hier auf eine eingehende Beschreibung dieser molekularen Vorgänge zu verzichten haben und dafür nur die Parameter aufsuchen, von denen der meßbare Verzerrungszustand der Körper an jeder Stelle abhängt. Alsdann wird festzustellen sein, wie die Form der Abhängigkeit der Lagranschen Funktion von diesen Parametern ist, die sich ja eigentlich aus kinetischer und potentieller Energie der einzelnen Molekel zusammensetzen wird. Ähnlich wird man in der Thermodynamik nicht auf die Schwingungen der Molekel eingehen, sondern die Temperatur selbst als allgemeinen Parameter einführen, und die Abhängigkeit der Energie von ihr untersuchen." (Hilbert 1906, 8–9)

tionism. The process becomes gradually manifest after 1910, though Hilbert still stuck to his original conceptions until around 1913.

The course on mechanics in the winter semester of 1910–1911 opened with an unambiguous statement about the essential role of mechanics as the foundation of natural science in general (Hilbert 1910–1911, 6). Hilbert praised the textbooks of Hertz and Boltzmann for their successful attempts to present in similar methodological terms, albeit starting from somewhat different premises, a fully axiomatic derivation of mechanics. This kind of presentation, Hilbert added, was currently being disputed. The course itself covered the standard topics of classical mechanics. Towards the end, however, Hilbert spoke about the "new mechanics." In this context he neither used the word "relativity" nor mentioned Einstein. Rather, he mentioned only Lorentz and spoke of invariance under the Lorentz transformations of all differential equations that describe natural phenomena as the main feature of this new mechanics. Hilbert stressed that the Newtonian equations of the "old" mechanics do not satisfy this basic principle, which, like Minkowski, he called the *Weltpostulat*. These equations must therefore be transformed, he said, so that they become Lorentz-invariant.[148] Hilbert showed that if the Lorentz transformations are used instead of the "Newton transformations," then the velocity of light is the same for every non-accelerated, moving system of reference.

Hilbert also mentioned the unsettled question of the status of gravitation in the framework of this new mechanics. He connected his presentation directly to Minkowski's sketchy treatment of this topic in 1909, and, like his friend, Hilbert does not seem to have been really bothered by the difficulties related with it. One should attempt to modify the Newtonian law in order to make it comply to the world-postulate, Hilbert said, but we must exercise special care when doing this since the Newtonian law has proved to be in the closest accordance with experience. As Hilbert knew from Minkowski's work, an adaptation of gravitation to the new mechanics would imply that its effects must propagate at the speed of light. This latter conclusion contradicts the "old theory," while in the framework of the "new mechanics," on the contrary, it finds a natural place. In order to adapt the Newtonian equations to the new mechanics, concluded Hilbert, we proceed, "as Minkowski did, via electromagnetism."[149]

The manuscript of the course does not record whether in the classroom Hilbert showed how, by proceeding "as Minkowski did, via electromagnetism," the adaptation of Newton's law should actually be realized. Perhaps at that time he still believed that Minkowski's early sketch could be further elaborated. Be that as it may, the concerns expressed here by Hilbert are not unlike those of other, contemporary physicists involved in investigating the actual place of the postulate of relativity in the general

148 "Alle grundlegenden Naturgesetzen entsprechenden Systeme von Differentialgleichungen sollen gegenüber der Lorentz-Transformation kovariant sein. ... Wir können durch Beobachtung von irgend welchen Naturvorgängen niemals entscheiden, ob wir ruhen, oder uns gleichförmig bewegen. Diesem Weltpostulate genügen die Newtonschen Gleichungen der älteren Mechanik nicht, wenn wir die Lorentz Transformation zugrunde legen: wir stehen daher vor der Aufgabe, sie dementsprechend umzugestalten." (Hilbert 1910–1911, 292)

picture of physics. It is relevant to recall at this stage, however, that Einstein himself published nothing on this topic between 1907 and June 1911.

9. KINETIC THEORY

After another standard course on continuum mechanics in the summer of 1911, Hilbert taught a course specifically devoted to kinetic theory of gases for the first time in the winter of 1911–1912. This course marked the starting point of Hilbert's definitive involvement with a broader range of physical theories. Hilbert opened the course by referring once again to three possible, alternative treatments of any physical theory. First, is the "phenomenological perspective,"[150] often applied to study the mechanics of continua. Under this perspective, the whole of physics is divided into various chapters, each of which can be approached using different, specific assumptions, from which different mathematical consequences can be derived. The main mathematical tool used in this approach is the theory of partial differential equations. In fact, much of what Hilbert had done in his 1905 lectures on the axiomatization of physics, and then in 1906 on mechanics of continua, could be said to fall within this approach.

The second approach that Hilbert mentioned assumes the validity of the "theory of atoms." In this case a "much deeper understanding is reached. ... We attempt to put forward a system of axioms which is valid for the whole of physics, and which enables all physical phenomena to be explained from a unified point of view."[151] The mathematical methods used here are obviously quite different from those of the phenomenological approach: they can be subsumed, generally speaking, under the methods of the theory of probabilities. The most salient examples of this approach are found in the theory of gases and in radiation theory. From the point of view of this approach, the phenomenological one is a palliative, indispensable as a primitive stage on the way to knowledge, which must however be abandoned "as soon as possible, in order to penetrate the real sanctuary of theoretical physics."[152] Unfortunately, Hilbert

149 "Wir können nun an die Umgestaltung des Newtonsches Gesetzes gehen, dabei müssen wir aber Vorsicht verfahren, denn das Newtonsche Gesetz ist das desjenige Naturgesetz, das durch die Erfahrung in Einklang bleiben wollen. Dieses wird uns gelingen, ja noch mehr, wir können verlangen, dass die Gravitation sich mit Lichtgeschwindigkeit fortpflanzt. Die alte Theorie kann das nicht, eine Fortpflanzung der Gravitation mit Lichtgeschwindigkeit widerspricht hier der Erfahrung: Die neue Theorie kann es, und man ist berechtigt, das als eine Vorzug derselben anzusehen, den eine momentane Fortpflanzung der Gravitation passt sehr wenig zu der modernen Physik. Um die Newtonschen Gleichungen für die neue Mechanik zu erhalten, gehen wir ähnlich vor wie Minkowski in der Elektromagnetik." (Hilbert 1910–1911, 295)

150 Boltzmann had used the term in this context in his 1899 Munich talk that Hilbert had attended. Cf. (Boltzmann 1899, 92–96).

151 "Hier ist das Bestreben, ein Axiomensystem zu schaffen, welches für die ganze Physik gilt, und aus diesem einheitlichen Gesichtspunkt alle Erscheinungen zu erklären. ... Jedenfalls gibt sie unvergleichlich tieferen Laufschuhes über Wesen und Zusammenhang der physikalischen Begriffe, ausserdem auch neue Aufklärung über physikalische Tatsachen, welche weit über die bei A) erhaltene hinausgeht." (Hilbert 1911–1912, 2)

said, mathematical analysis is not yet developed sufficiently to provide for all the demands of the second approach. One must therefore do without rigorous logical deductions and be temporarily satisfied with rather vague mathematical formulae.[153] Hilbert considered it remarkable that by using this method one nevertheless obtains ever new results that are in accordance with experience. He thus declared that the "main task of physics," embodied in the third possible approach, would be "the molecular theory of matter" itself, standing above the kinetic theory, as far as its degree of mathematical sophistication and exactitude is concerned. In the present course, Hilbert intended to concentrate on kinetic theory, yet he promised to consider the molecular theory of matter in the following semester. He did so, indeed, a year later.

Many of the important innovations implied by Hilbert's solution of the Boltzmann equation are already contained in this course of 1911–1912.[154] It was Maxwell in 1860 who first formulated an equation describing the distribution of the number of molecules of a gas, with given energy at a given point in time. Maxwell, however, was able to find only a partial solution which was valid only for a very special case.[155] In 1872 Boltzmann reformulated Maxwell's equation in terms of a single, rather complex, integro-differential equation, that has remained associated with his name ever since. The only exact solution Boltzmann had been able to find, however, was still valid for the same particular case that Maxwell had treated in his own model (Boltzmann 1872). By 1911, some progress had been made on the solution of the Boltzmann equation. The laws obtained from the partial knowledge concerning those solutions, which described the macroscopic movement and thermal processes in gases, seemed to be qualitatively correct. However, the mathematical methods used in the derivations seemed inconclusive and sometimes arbitrary. It was quite usual to rely on average magnitudes and thus the calculated values of the coefficients of heat conduction and friction appeared to be dubious. A more accurate estimation of these values remained a main concern of the theory, and the techniques developed by Hilbert apparently offered the means to deal with it.[156]

Very much as he had done with other theories in the past, Hilbert wanted to show how the whole kinetic theory could be developed starting from one basic formula, which in this case would be precisely the Boltzmann equation. His presentation would depart from the phenomenological approach by making some specific assumptions about the molecules, namely that they are spheres identical to one another in

152 "Wenn man auf diesem Standpunkt steht, so wird man den früheren nur als einer Notbehelf bezeichnen, der nötig ist als eine erste Stufe der Erkenntnis, über die man aber eilig hinwegschreiten muss, um in die eigentlichen Heiligtümer der theoretischen Physik einzudringen." (Hilbert 1911–1912, 2)

153 "... sich mit etwas verschwommenen mathematischen Formulierungen zufrieden geben muss." (Hilbert 1911–1912, 2)

154 In fact, in December 1911 Hilbert presented to the *Göttinger Mathematische Gesellschaft* an overview of his recent investigations on the theory, stating that he intended to publish them soon. Cf. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 21 (1912), 58.

155 Cf. (Brush 1976, 432–446).

156 Cf. (Born 1922, 587–589).

size. In addition he would focus, not on the velocity of any individual such molecule, but rather on their velocity distribution φ over a small element of volume.

In the opening lectures of the course, a rather straightforward discussion of the elementary physical properties of a gas led Hilbert to formulate a quite complicated equation involving φ. Hilbert asserted that a general solution of this equation was impossible, and it was thus necessary to limit the discussion to certain specific cases (Hilbert 1911–1912, 21). In the following lectures he added some specific, physical assumptions concerning the initial and boundary conditions for the velocity distribution in order to be able to derive more directly solvable equations. These assumptions, which he formulated as axioms of the theory, restricted the generality of the problem to a certain extent, but allowed for representing the distribution function as a series of powers of a certain parameter. In a first approximation, the relations between the velocity distributions yielded the Boltzmann distribution. In a second approximation, they yielded the propagation of the average velocities in space and time. Under this representation the equation appeared as a linear symmetric equation of the second type, where the velocity distribution φ is the unknown function, thus allowing the application of Hilbert's newly developed techniques. Still, he did not prove in detail the convergence of the power series so defined, nor did he complete the evaluation of the transport coefficient appearing in the distribution formula.

Hilbert was evidently satisfied with his achievement in kinetic theory. He was very explicit in claiming that without a direct application of the techniques he had developed in the theory of integral equations, and without having formulated the physical theory in terms of such integral equations, it would be impossible to provide a solid and systematic foundation for the theory of gases as currently known (Hilbert 1912a, 268; 1912b, 562). And very much as with his more purely mathematical works, also here Hilbert was after a larger picture, searching for the underlying connections among apparently distant fields. Particularly interesting for him were the multiple connections with radiation theory, which he explicitly mentioned at the end of his 1912 article, thus opening the way for his forthcoming courses and publications. In his first publication on radiation theory he explained in greater detail and with unconcealed effusiveness the nature of this underlying connection. He thus said:

> In my treatise on the "Foundations of the kinetic theory of gases," I have shown, using the theory of linear integral equations, that starting alone from the Maxwell-Boltzmann fundamental formula —the so-called collision formula— it is possible to construct systematically the kinetic theory of gases. This construction is such, that it requires only a consistent implementation of the methods of certain mathematical operations prescribed in advance, in order to obtain the proof of the second law of thermodynamics, of Boltzmann's expression for the entropy of a gas, of the equations of motion that take into account both the internal friction and the heat conduction, and of the theory of diffusion of several gases. Likewise, by further developing the theory, we obtain the precise conditions under which the law of equipartition of energies over the intermolecular parameter is valid. Concerning the motion of compound molecules, a new law is also obtained according to which the continuity equation of hydrodynamics has a much more general meaning than the usual one. ...

> Meanwhile, there is a second physical domain whose principles have not yet been investigated at all from the mathematical point of view, and for the establishment of whose foundations—as I have recently discovered—the same mathematical tools provided by the integral equations are absolutely necessary. I mean by this the elementary theory of radiation, understanding by it the phenomenological aspect of the theory, which at the most immediate level concerns the phenomena of emission and absorption, and on top of which stand Kirchhoff's laws concerning the relations between emission and absorption. (Hilbert 1912b, 217–218)

Hilbert could boast now two powerful mathematical tools that allowed him to address the study of a broad spectrum of physical theories. On the one hand, the axiomatic method would help dispel conceptual difficulties affecting established theories—thus fostering their continued development—and also open the way for a healthy establishment of new ones. In his earlier courses he had already explored examples of the value of the method for a wide variety of disciplines, but Minkowski's contributions to electrodynamics and his analysis of the role of the principle of relativity offered perhaps, from Hilbert's point of view, the most significant example so far of the actual realization of its potential contribution. On the other hand, the theory of linear integral equations had just proven its value in the solution of such a central, open problem of physics. As far as he could see from his own, idiosyncratic perspective, the program for closing the gap between physical theories and mathematics had been more successful so far than he may have actually conceived when posing his sixth problem back in 1900. Hilbert was now prepared to attack yet another central field of physics and he would do so by combining once again the two mathematical components of his approach. The actual realization of this plan, however, was less smooth than one could guess from the above-quoted, somewhat pompous, declaration. As will be seen in the next section, although Hilbert's next incursion into the physicists' camp led to some local successes, as a whole they were less impressive in their overall significance than Hilbert would have hoped.

But even though Hilbert was satisfied with what his mastery of integral equations had allowed him to do thus far, and with what his usual optimism promised to achieve in other physical domains in the near future, there was an underlying fundamental uneasiness that he was not able to conceal behind the complex integral formulas and he preferred to explicitly share this uneasiness with his students. It concerned the possible justification of using probabilistic methods in physics in general and in kinetic theory in particular. Hilbert's qualms are worth quoting in some detail:

> If Boltzmann proves … that the Maxwell distribution … is the most probable one from among all distributions for a given amount of energy, this theorem possesses in itself a certain degree of interest, but it does not allow even a minimal inference concerning the velocity distribution that actually occurs in any given gas. In order to lay bare the core of this question, I want to recount the following example: in a raffle with one winner out of 1000 tickets, we distribute 998 tickets among 998 persons and the remaining two we give to a single person. This person thus has the greatest chance to win, compared to all other participants. His probability of winning is the greatest, and yet it is highly improbable that he will win. The probability of this is close to zero. In the same fashion, the probability of occurrence of the Maxwell velocity distribution is greater than that of any other

> distribution, but equally close to zero, and it is therefore almost absolutely certain that the Maxwell distribution will not occur.
>
> What is needed for the theory of gases is much more than that. We would like to prove that for a specified distribution, there is a probability very close to 1 that distribution is asymptotically approached as the number of molecules becomes infinitely large. And in order to achieve that, it is necessary to modify the concept of "velocity distribution" in order to obtain some margin for looseness. We should formulate the question in terms such as these: What is the probability for the occurrence of a velocity distribution that deviates from Maxwell's by no more than a given amount? And moreover: what allowed deviation must we choose in order to obtain the probability 1 in the limit?[157]

Hilbert discussed in some detail additional difficulties that arise in applying probabilistic reasoning within kinetic theory. He also gave a rough sketch of the kind of mathematical considerations that could in principle provide a way out to the dilemmas indicated. Yet he made clear that he could not give final answers in this regard.[158] This problem would continue to bother him in the near future. In any case, after this brief excursus, Hilbert continued with the discussion he had started in the first part of his lectures and went on to generalize the solutions already obtained to the cases of mixtures of gases or of polyatomic gases.

In spite of its very high level of technical sophistication of his approach to kinetic theory, it is clear that Hilbert did not want his contribution to be seen as a purely mathematical, if major, addition to the solution of just one central, open problem of

---

157 "Wenn z.B. Boltzmann beweist—übrigens auch mit einigen Vernachlässigungen—dass die Maxwellsche Verteilung (die nach dem Exponentialgesetz) unter allen Verteilungen von gegebener Gesamtenergie die wahrscheinlichste ist, so besitzt dieser Satz ja an und für sich ein gewisses Interesse, aber er gestattet auch nicht der geringsten Schluss auf die Geschwindigkeitsverteilung, welche in einem bestimmten Gase wirklich eintritt. Um den Kernpunkt der Frage klar zu legen, will ich an folgendes Beispiel erinnern: In einer Lotterie mit einem Gewinn und von 1000 Losen seien 998 Losen auf 998 Personen verteilt, die zwei übrigen Lose möge eine andere Person erhalten. Dann hat diese Person im Vergleich zu jeder einzelnen andern die grössten Gewinnchancen. Die Wahrscheinlichkeit des Gewinnen ist für sie am grössten, aber es ist immer noch höchst unwahrscheinlich, dass sie gewinnt. Denn die Wahrscheinlichkeit ist so gut wie Null.
Ganz ebenso ist die Wahrscheinlichkeit für den Eintritt der Maxwellschen Geschwindigkeitsverteilung zwar grösser als die für das Eintreten einer jeden bestimmten andern, aber doch noch so gut wie Null, und es ist daher fast mit absoluter Gewissheit sicher, dass die Maxwellsche Verteilung nicht eintritt.
Was wir für die Gastheorie brauchen, ist sehr viel mehr. Wir wünschen zu beweisen, dass für eine gewisse ausgezeichnete Verteilung eine Wahrscheinlichkeit sehr nahe an 1 besteht, derart, dass sie sich mit Unendliche wachsende Molekülzahl der 1 asymptotisch annähert. Und um das zu erreichen, müssen wir den Begriff der „Geschwindigkeitsverteilung" etwas modifizieren, indem wir einen gewissen Spielraum zulassen. Wir hätten die Frage etwa so zu formulieren: Wie gross ist die Wahrscheinlichkeit dafür, dass eine Geschwindigkeitsverteilung eintritt, welche von der Maxwellschen nur um höchstens einen bestimmten Betrag abweicht—und weiter: wie gross müssen wir die zugelassenen Abweichungen wählen, damit wir im limes die Wahrscheinlichkeit eins erhalten?" (Hilbert 1911–1912, 75–76)

158 "Ich will Ihnen nun auseinandersetzen, wie ich mir etwa die Behandlung dieser Frage denke. Es sind da sicher noch grosse Schwierigkeiten zu überwinden, aber die Idee nach wird man wohl in folgender Weise vorgehen müssen: ... " (Hilbert 1911–1912, 77)

this theory. Rather, his aim was to be directly in touch with the physical core of this and other, related domains. The actual scope of his physical interests at the time becomes more clearly evident in a seminar that he organized in collaboration with Erich Hecke (1887–1947), shortly after the publication of his article on kinetic theory.[159] The seminar was also attended by the Göttingen docents Max Born, Paul Hertz, Theodor von Kármán (1881–1963), and Erwin Madelung (1881–1972), and the issues discussed included the following:[160]

- the ergodic hypothesis and its consequences;
- on Brownian motion and its theories;
- electron theory of metals in analogy to Hilbert's theory of gases;
- report on Hilbert's theory of gases;
- on dilute gases;
- theory of dilute gases using Hilbert's theory;
- on the theory of chemical equilibrium, including a reference to the
- related work of Sackur;
- dilute solutions.

The names of the participants and younger colleagues indicate that these deep physical issues, related indeed with kinetic theory but mostly not with its purely mathematical aspects, could not have been discussed only superficially. Especially indicative of Hilbert's surprisingly broad spectrum of interests is the reference to the work of Otto Sackur (1880–1914). Sackur was a physical chemist from Breslau whose work dealt mainly with the laws of chemical equilibrium in ideal gases and on Nernst law of heat. He also wrote a widely used textbook on thermochemistry and thermodynamics (Sackur 1912). His experimental work was also of considerable significance and, more generally, his work was far from the typical kind of purely technical, formal mathematical physics that is sometimes associated with Hilbert and the Göttingen school.[161]

10. RADIATION THEORY

Already in his 1911–1912 lectures on kinetic theory, Hilbert had made clear his interest in investigating, together with this domain and following a similar approach, the

159 Hecke had also taken the notes of the 1911–1912 course.

160 References to this seminar appear in (Lorey 1916, 129). Lorey took this information from the German student's journal *Semesterberichte des Mathematischen Vereins*. The exact date of the seminar, however, is not explicitly stated.

161 See Sackur's obituary in *Physikalische Zeitschrift* 16 (1915), 113–115. According to Reid's account (1970, 129), Ewald succinctly described Hilbert's scientific program at the time of his arrival in Göttingen with the following, alleged quotation of the latter: "We have reformed mathematics, the next thing to reform is physics, and then we'll go on to chemistry." Interest in Sackur's work, as instantiated in this seminar would be an example of an intended, prospective attack on this field. There are not, however, many documented, further instances of this kind.

theory of radiation.[162] Kirchhoff's laws of emission and absorption had traditionally stood as the focus of interest of this theory. These laws, originally formulated in late 1859, describe the energetic relations of radiation in a state of thermal equilibrium.[163] They assert that in the case of purely thermal radiation (i.e., radiation produced by thermal excitation of the molecules) the ratio between the emission and absorption capacities of matter, η and α respectively, is a universal function of the temperature T and the wavelength λ,

$$\frac{\eta}{\alpha} = K(T, \lambda)$$

and is therefore independent of the substance and of any other characteristics of the body in question. One special case that Kirchhoff considered in his investigations is the case $\alpha = 1$, which defines a "black body," namely, a hypothetical entity that completely absorbs all wavelengths of thermal radiation incident in it.[164]

In the original conception of Kirchhoff's theory the study of black-body radiation may not have appeared as its most important open problem, but in retrospect it turned out to have the farthest-reaching implications for the development of physics at large. In its initial phases, several physicists attempted to determine over the last decades of the century the exact form of the spectral distribution of the radiation $K(T, \lambda)$ for a black body. Prominent among them was Wilhelm Wien, who approached the problem by treating this kind of radiation as loosely analogous to gas molecules. In 1896 he formulated a law of radiation that predicted very accurately recent existing measurements. Planck, however, was dissatisfied with the lack of a theoretical justification for what seemed to be an empirically correct law. In searching for such a justification within classical electromagnetism and thermodynamics, he modeled the atoms at the inside walls of a black-body cavity as a collection of electrical oscillators which absorbed and emitted energy at all frequencies. In 1899 he came forward with an expression for the entropy of an ideal oscillator, built on an analogy with Boltzmann's kinetic theory of gases, that provided the desired theoretical justification of Wien's law (Planck 1899). Later on, however, additional experiments produced values for the spectrum at very low temperatures and at long wavelengths that were not anymore in agreement with this law.

Another classical attempt was advanced by John William Strutt, Lord Rayleigh (1842–1919), and James Jeans (1877–1946), also at the beginning of the century.[165] Considering the radiation within the black-body cavity to be made up of a series of standing waves, they derived a law that, contrary to Wien's, approximated experimen-

162 Minkowski and Hilbert even had planned to have a seminar on the theory of heat radiation as early as 1907 (Minkowski 1907).

163 Cf., e.g., (Kirchhoff 1860).

164 Cf. (Kuhn 1978, 3–10).

165 Cf. (Kuhn 1978, 144–152).

tal data very well at long wavelengths but failed at short ones. In the latter case, it predicted that the spectrum would rise to infinity as the wavelength decreased to zero.[166]

In a seminal paper of 1900, Planck formulated an improved law that approximated Wien's formula in the case of short wavelengths and the Rayleigh-Jeans law in the case of long wavelengths. The law assumed that the resonator entropy is calculated by counting the number of distributions of a given number of finite, equal "energy elements" over a set of resonators, according to the formula:

$$E = nhv$$

where n is an integer, v is the oscillators' frequency, and h is the now famous Planck constant, h = 6.55 x 10 -27 erg-sec. (Planck 1900). Based on this introduction of energy elements, assuming thermal equilibrium and applying statistical methods of kinetic theory, Planck derived the law that he had previously obtained empirically and that described the radiant energy distribution of the oscillators:

$$U_v = \frac{hv}{e^{hv/kT} - 1}.$$

Planck saw his assumption of energy elements as a convenient mathematical hypothesis, and not as a truly physical claim about the way in which matter and radiation actually interchange energy. In particular, he did not stress the significance of the finite energy elements that entered his calculation and he continued to think about the resonators in terms of a continuous dynamics. He considered his assumption to be very important since it led with high accuracy to a law that had been repeatedly confirmed at the experimental level, but at the same time he considered it to be a provisional one that would be removed in future formulations of the theory. In spite of its eventual revolutionary implications on the developments of physics, Planck did not realize before 1908 that his assumptions entailed any significant departure from the fundamental conceptions embodied in classical physics. As a matter of fact, he did not publish any further research on black-body radiation between 1901 and 1906.[167]

The fundamental idea of the quantum discontinuity was only slowly absorbed into physics, first through the works of younger physicists such as Einstein, Laue and Ehrenfest, then by leading ones such as Planck, Wien and Lorentz, and finally by their readers and followers. The details pertaining to this complex process are well beyond the scope of my account here. Nonetheless, it is worth mentioning that a very significant factor influencing Planck's own views in this regard was his correspondence with Lorentz in 1908. Lorentz had followed with interest since 1901 the debates around black-body radiation, and he made some effort to connect them with his own theory of the electron. At the International Congress of Mathematicians held

166 Much later Ehrenfest (1911) dubbed this phenomenon "ultraviolet catastrophe."

167 This is the main claim developed in detail in the now classic (Kuhn 1978). For a more recent, summary account of the rise of quantum theory, see (Kragh 1999, chap. 5).

in Rome in 1908, Lorentz was invited to deliver one of the plenary talks, which he devoted to this topic. This lecture was widely circulated and read thereafter and it represented one of the last attempts at interpreting cavity radiation in terms of a classical approach (Lorentz 1909). But then, following critical remarks by several colleagues, Lorentz added a note to the printed version of his talk where he acknowledged that his attempt to derive the old Rayleigh-Jeans radiation law from electron theory was impracticable unless the foundations of the latter would be deeply modified. A letter to Lorentz sent by Planck in the aftermath of the publication contains what may be the latter's first acknowledgment of the need to introduce discontinuity as a fundamental assumption. Lorentz himself, at any rate, now unambiguously adopted the idea of energy quanta and he stressed it explicitly in his lectures of early 1909 in Utrecht.[168] Later, in his 1910 Wolfskehl cycle in Göttingen, Lorentz devoted one of the lectures to explaining why the classical Hamilton principle would not work for radiation theory. An "entirely new hypothesis," he said, needed to be introduced. The new hypothesis he had in mind was "the introduction of the energy elements invented by Planck" (Lorentz 1910, 1248). Hilbert was of course in the audience and he must have attentively listened to his guest explaining the innovation implied by this fundamental assertion.

Starting in 1911 research on black-body radiation became less and less prominent and at the same time the quantum discontinuity hypothesis became a central issue in other domains such as thermodynamics, specific heats, x-rays, and atomic models. The apparent conflicts between classical physics and the consequences of the hypothesis stood at the focus of discussions in the first Solvay conference organized in Brussels in 1911.[169] These discussion prompted Poincaré, who until then was reticent to adopt the discontinuity hypothesis, to elaborate a mathematical proof that Planck's radiation law necessarily required the introduction of quanta (Poincaré 1912). His proof also succeeded in convincing Jeans in 1913, who thus became one of the latest prominent physicists to abandon the classical conception in favor of discontinuity (Jeans 1914).[170]

The notes of Hilbert's course on radiation theory in the summer semester of 1912, starting in late April, evince a clear understanding and a very broad knowledge of all the main issues of the discipline. In his previous course on kinetic theory, Hilbert had promised to address "the main task of physics," namely, the molecular theory of matter itself, a theory he described as having a greater degree of mathematical sophistication and exactitude than kinetic theory. To a certain extent, teaching this course meant fulfilling that promise. Hilbert declared that he intended to address now the "domain of physics properly said," which is based on the point of view of the atomic theory. Hilbert was clearly very much impressed by recent developments in quantum theory. "Never has there been a more proper and challenging time than now," he said, "to

168 Cf. (Kuhn 1978, 189–197).
169 Cf. (Barkan 1993).
170 Cf. (Kuhn 1978, 206–232).

undertake the research of the foundations of physics." What seems to have impressed Hilbert more than anything else were the deep interconnections recently discovered in physics, "of which formerly no one could have even dreamed, namely, that optics is nothing but a chapter within the theory of electricity, that electrodynamics and thermodynamics are one and the same, that energy also possesses inertial properties, that physical methods have been introduced into chemistry as well."[171] And above all, the "atomic theory," the "principle of discontinuity," which was not a hypothesis anymore, but rather, "like Copernicus's theory, a fact confirmed by experiment."[172]

Hilbert opened with a summary account of four-vector analysis[173] and of Special Theory of Relativity. Taking the relativity postulate to stand "on top" of physics as a whole, he then formulated the basics of electrodynamics as currently conceived, including Born's concept of a rigid body. This is perhaps Hilbert's first systematic discussion of Special Theory of Relativity in his lecture courses. As in the case of kinetic theory, Hilbert already raised here some of the ideas that he would later develop in his related, published works. But again, the course was far from being just an exercise in applying integral equations techniques to a particularly interesting, physical case. Rather, Hilbert covered most of the core, directly relevant, physical questions. Thus, among the topics discussed in the course we find the energy distribution of black-body radiation (including a discussion of Wien's and Rayleigh's laws) and Planck's theory of resonators under the effect of radiation. Hilbert particularly stressed the significance of recent works by Ehrenfest and Poincaré, as having shown the necessity of a discontinuous form of energy distribution (Hilbert 1912c, 94).[174] Hilbert also made special efforts to have Sommerfeld invited to give the last two lectures in the course, in which important, recent topics in the theory were discussed.[175]

However, as with all other physical theories, what Hilbert considered to be the main issue of the theory of radiation as a whole was the determination of the precise form of a specific law that stood at its core. In this case the law in question was Kirchhoff's law of emission and absorption, to which Hilbert devoted several lectures. Of particular interest for him was the possibility of using the techniques of the

171 "Nun kommen wir aber zu eigentlicher Physik, welche sich auf der Standpunkt der Atomistik stellt und da kann man sagen, dass keine Zeit günstiger ist und keine mehr dazu herausfordert, die Grundlagen dieser Disziplin zu untersuchen, wie die heutige. Zunächst wegen der Zusammenhänge, die man heute in der Physik entdeckt hat, wovon man sich früher nichts hätte träumen lassen, dass die Optik nur ein Kapitel der Elektrizitätslehre ist, dass Elektrodynamik und Thermodynamik dasselbe sind, dass auch die Energie Trägheit besitzt, dann dass auch in der Chemie (Metalchemie, Radioaktivität) physikalische Methoden in der Vordergrund haben." (Hilbert 1912c, 2)

172 "... wie die Lehre des Kopernikus, eine durch das Experimente bewiesene Tatsache." (Hilbert 1912c, 2)

173 A hand-written addition to the typescript (Hilbert 1912c, 4) gives here a cross-reference to Hilbert's later course, (Hilbert 1916, 45–56), where the same topic is discussed in greater detail.

174 He referred to (Ehrenfest 1911) and (Poincaré 1912). Hilbert had recently asked Poincaré for a reprint of his article. See Hilbert to Poincaré, 6 May 1912. (Hilbert 1932–1935, 546)

175 Cf. Hilbert to Sommerfeld, 5 April 1912 (*Nachlass* Arnold Sommerfeld, Deutsches Museum, Munich. HS1977–28/A, 141).

theory of integral equations for studying the foundations of the law and providing a complete mathematical justification for it. This would also become the main task pursued in his published articles on the topic, which I discuss in detail in the next four sections. In fact, just as his summer semester course was coming to a conclusion, Hilbert submitted for publication his first paper on the "Foundations of the Elementary Theory of Radiation."

11. STRUCTURE OF MATTER AND RELATIVITY: 1912–1914

After this account of Hilbert's involvement with kinetic theory and radiation theory, I return to 1912 in order to examine his courses in physics during the next two years.[176] The structure of matter was the focus of attention here, and Hilbert now finally came to adopt electromagnetism as the fundamental kind of phenomenon to which all others should be reduced. The atomistic hypothesis was a main physical assumption underlying all of Hilbert's work from very early on, and also in the period that started in 1910. This hypothesis, however, was for him secondary to more basic, mathematical considerations of simplicity and precision. A main justification for the belief in the validity of the hypothesis was the prospect that it would provide a more accurate and detailed explanation of natural phenomena once the tools were developed for a comprehensive mathematical treatment of theories based on it. Already in his 1905 lectures on the axiomatization of physics, Hilbert had stressed the problems implied by the combined application of analysis and the calculus of probabilities as the basis for the kinetic theory, an application that is not fully justified on mathematical grounds. In his physical courses after 1910, as we have seen, he again expressed similar concerns. The more Hilbert became involved with the study of kinetic theory itself, and at the same time with the deep mathematical intricacies of the theory of linear integral equations, the more these concerns increased. This situation, together with his growing mastery of specific physical issues from diverse disciplines, helps to explain Hilbert's mounting interest in questions related to the structure of matter that occupied him in the period I discuss now. The courses described below cover a wide range of interesting physical questions. In this account, for reasons of space, I will comment only on those aspects that are more directly connected with the questions of axiomatization, reductionism and the structure of matter.

11.1 Molecular Theory of Matter - 1912–1913

Hilbert's physics course in the winter semester of 1912–1913 was devoted to describing the current state of development of the molecular theory of matter (Hilbert 1912–

176 The printed version of the *Verzeichnis der Vorlesungen an der Georg-August-Universität zu Göttingen* registers several courses for which no notes or similar documents are extant, and about which I can say nothing here: summer semester, 1912 - Mathematical Foundations of Physics; winter semester, 1912–1913 - Mathematical Foundations of Physics.

1913),[177] and particularly the behavior of systems of huge quantities of particles moving in space, and affecting each other through collisions and other kinds of interacting forces.[178] The first of the course's three parts deals with the equation of state, including a section on the principles of statistical mechanics. The second part is characterized as "phenomenological" and the third part as "kinetic," in which entropy and the quantum hypothesis are discussed. This third part also includes a list of axioms for the molecular theory of matter. Hilbert was thus closing a circle initiated with the course on kinetic theory taught one year earlier.

Hilbert suggested that the correct way to come to terms with the increasingly deep mathematical difficulties implied by the atomistic hypothesis would be to adopt a "physical point of view." This means that one should make clear, through the use of the axiomatic method, those places in which physics intervenes into mathematical deduction. This would allow separating three different components in any specific physical domain considered: first, what is arbitrarily adopted as definition or taken as an assumption of experience; second, what a-priori expectations follow from these assumptions, which the current state of mathematics does not yet allow us to conclude with certainty; and third, what is truly proven from a mathematical point of view.[179] This separation interestingly brings to mind Minkowski's earlier discussion on the status of the principle of relativity. It also reflects to a large extent the various levels of discussion evident in Hilbert's articles on radiation theory, and it will resurface in his reconsideration of the view of mechanics as the ultimate explanation of physical phenomena.

In the first part of the course, Hilbert deduced the relations between pressure, volume and temperature for a completely homogenous body. He considered the body as a mechanical system composed of molecules, and applied to it the standard laws of mechanics. This is a relatively simple case, he said, that can be easily and thoroughly elucidated. However, deriving the state equation and explaining the phenomenon of condensation covers only a very reduced portion of the empirically manifest properties of matter. Thus the second part of the lectures was devoted to presenting certain, more complex physical and chemical phenomena, the kinetic significance of which would then be explained in the third part of the course.[180] The underlying approach

177 A second copy of the typed notes in found in *Nachlass* Max Born, Staatsbibliothek Berlin, Preussischer Kulturbesitz #1817.

178 "Das Ziel der Vorlesung ist es, die Molekulartheorie der Materie nach dem heutigen Stande unseres Wissens zu entwickeln. Diese Theorie betrachtet die physikalischen Körper und ihre Veränderungen unter dem Scheinbilde eines Systems ungeheuer vieler im Raum bewegter Massen, die durch die Stösse oder durch andere zwischen ihnen wirkenden Kräfte einander beeinflussen." (Hilbert 1912–1913, 1)

179 "Dabei werden wir aber streng axiomatisch die Stellen, in denen die Physik in die mathematische Deduktion eingreift, deutlich hervorheben, und das voneinander trennen, was erstens als logisch willkürliche Definition oder Annahme der Erfahrung entnommen wird, zweitens das, was a priori sich aus diesen Annahmen folgern liesse, aber wegen mathematischer Schwierigkeiten zur Zeit noch nicht sicher gefolgert werden kann, und dritten, das, was bewiesene mathematische Folgerung ist." (Hilbert 1912–1913, 1)

was to express the basic facts of experience in mathematical language, taking them as axioms in need of no further justification. Starting from these axioms one would then deduce as many results as possible, and the logical interdependence of these axioms would also be investigated. In this way, Hilbert declared, the axiomatic method, long applied in mathematics with great success, can also be introduced into physics.[181]

A main task that Hilbert had pursued in his 1905 lectures on axiomatization was to derive, from general physical and mathematical principles in conjunction with the specific axioms of the domain in question, an equation that stands at the center of each discipline and that accounts for the special properties of the particular system under study. Hilbert explicitly stated this as a main task for his system of axioms also in the present case.[182] A first, general axiom he introduced was the "principle of equilibrium," which reads as follows:

> In a state of equilibrium, the masses of the independent components are so distributed with respect to the individual interactions and with respect to the phases, that the characteristic function that expresses the properties of the system attains a minimum value.[183]

Hilbert declared that such an axiom had not been explicitly formulated before and claimed that its derivation from mechanical principles should be done in terms of purely kinetic considerations, such as would be addressed in the third part of the course.[184] At the same time he stated that, in principle, this axiom is equivalent to the second law of thermodynamics, which Hilbert had usually formulated in the past as

---

180 "Wir haben bisher das Problem behandelt, die Beziehung zwischen $p, v,$ und ϑ an einem chemisch völlig homogenen Körper zu ermitteln. Unser Ziel war dabei, diese Beziehung nach den Gesetzen der Mechanik aus der Vorstellung abzuleiten, dass der Körper ein mechanisches System seiner Molekele ist. In dem bisher behandelten, besonders einfache Falle, in dem wir es mit einer einzigen Molekel zu tun hatten, liess sich dies Ziel mit einer gewissen Vollständigkeit erreichen. Eine in einem bestimmten Temperaturintervall mit der Erfahrung übereinstimmende Zustandsgleichung geht nämlich aus der Kinetischen Betrachtung hervor. Mit der Kenntnis der Zustandsgleichung und der Kondensationserscheinungen ist aber nur ein sehr kleiner Teil, der sich empirisch darbietenden Eigenschaften der Stoffe erledigt. *Wir werden daher in diesem zweiten Teile diejenigen Ergebnisse der Physik und Chemie zusammenstellen, deren kinetische Deutung wir uns später zur Aufgabe machen wollen*." (Hilbert 1912–1913, 50)

181 "Die reinen Erfahrungstatsachen werden dabei in mathematischer Sprache erscheinen und als Axiome auftreten, die hier keiner weiteren Begründung bedürfen. Aus diesen Axiomen werden wir soviel als möglich, rein mathematische Folgerungen ziehen, und dabei untersuchen, welche unter den Axiomen voneinander unabhängig sind und welche zum Teil auseinander abgeleitet werden können. Wir werden also den axiomatischen Standpunkt, der in der modernen Mathematik schon zur Geltung gebracht ist, auf die Physik anwenden." (Hilbert 1912–1913, 50)

182 "Um im einzelnen Falle die charakteristische Funktion in ihrer Abhängigkeit von der eigentlichen Veränderlichen und den Massen der unabhängigen Bestandteile zu ermitteln, müssen verschiedenen neue Axiome hinzugezogen werden." (Hilbert 1912–1913, 60)

183 "Im Gleichgewicht verteilen sich die Massen der unabhängigen Bestandteile so auf die einzelnen Verbindungen und Phasen, dass die charakteristische Funktion, die den Bedingungen des Systems entspricht, ein Minimum wird." (Hilbert 1912–1913, 60)

184 "Es muss kinetischen Betrachtung überlassen bleiben, es aus den Prinzipien der Mechanik abzuleiten und wir werden im dritten Teil der Vorlesung die erste Ansätze an solchen kinetischen Theorie kennen lernen." (Hilbert 1912–1913, 61)

the impossibility of the existence of a "*perpetuum mobile*."

The topics for which Hilbert carried out an axiomatic analysis included the equation of state and the third law of thermodynamics. Hilbert's three axioms for the former allowed him a derivation of the expression for the thermodynamical potential of a mixture of gases that was followed by a discussion of the specific role of each of the axioms involved.[185] Concerning the third law of thermodynamics, Hilbert introduced five axioms meant to account for the relationship between the absolute zero temperature, specific heats and entropy. Also in this case he devoted some time to discussing the logical and physical interdependence of these axioms. Hilbert explained that the axiomatic reduction of the most important theorems into independent components (the axioms) is nevertheless not yet complete. The relevant literature, he said, is also full of mistakes, and the real reason for this lies at a much deeper layer. The basic concepts seem to be defined unclearly even in the best of books. The problematic use of the basic concepts of thermodynamics went back in some cases even as far as Helmholtz.[186]

The third part of the course contained, as promised, a "kinetic" section especially focusing on a discussion of rigid bodies. Hilbert explained that the results obtained in the previous sections had been derived from experience and then generalized by means of mathematical formulae. In order to derive them a-priori from purely mechanical considerations, however, one should have recourse to the "fundamental principle of statistical mechanics,"[187] presumably referring to the assumption that all accessible states of a system are equally probable. Hilbert thought that the task of the course would be satisfactorily achieved if those results that he had set out to derive were indeed reduced to the theorems of mechanics together with this principle.[188] At any rate, the issues he discussed in this section included entropy, thermodynamics laws and the quantum hypothesis.

---

185 "Die drei gegebenen Axiome reichen also hin, um das thermodynamische Potential der Mischung zu berechnen. Aber sind nicht in vollem Umfange dazu Notwendig. Nimmt man z.B. das dritte Axiom für eine bestimmte Temperatur gültig an, so folgt es für jede beliebige Temperatur aus den beiden ersten Axiomen. Ebensowenig ist das erste und zweite Axiom vollständig voneinander unabhängig." (Hilbert 1912–1913, 66)

186 "Die axiomatische Reduktion der vorstehenden Sätze auf ihre unabhängigen Bestandteile ist demnach noch nicht vollständig durchgeführt, und es finden sich auch in der Literatur hierüber verschiedene Ungenauigkeiten. Was den eigentlichen Kern solcher Missverständnisse anlangt, so glaube ich, dass er tiefer liegt. Die Grundbegriffe scheinen mir selbst in den besten Lehrbüchern nicht genügend klar dargestellt zu sein, ja, in einem gleich zu erörternden Punkte geht die nicht ganz einwandfreie Anwendung der thermodynamischen Grundbegriffe sogar auch Helmholtz zurück." (Hilbert 1912–1913, 80)

187 "Um die empirisch gegebenen und zu mathematischen Formeln verallgemeinerten Ergebnisse des vorigen Teiles a priori und zwar auf rein mechanischem Wege abzuleiten, greifen wir wieder auf des Grundprinzip des statistischen Mechanik zurück, von der wir bereits im ersten Teil ausgegangen waren." (Hilbert 1912–1913, 88)

188 "Auf die Kritik dieses Grundprinzipes und die Grenzen, die seiner Anwendbarkeit gesteckt sind, können wir hier nicht eingehen. Wir betrachten vielmehr unser Ziel als erreicht, wenn die Ergebnisse, die abzuleiten wir uns zur Aufgabe stellen, auf die Sätze der Mechanik und auf jenes Prinzip zurückgeführt sind." (Hilbert 1912–1913, 88)

It is noteworthy that, although in December 1912, Born himself lectured on Mie's theory of matter at the *Göttinger Mathematische Gesellschaft*,[189] and that this theory touched upon many of the issues taught by Hilbert in this course, neither Mie's name nor his theory are mentioned in the notes. Nor was the theory of relativity mentioned in any way.

11.2 Electron Theory: 1913

In April of 1913 Hilbert organized a new series of Wolfskehl lectures on the current state of research in kinetic theory, to which he invited the leading physicists of the time. Planck lectured on the significance of the quantum hypothesis for kinetic theory. Peter Debye (1884–1966), who would become professor of physics in Göttingen the next year, dealt with the equation of state, the quantum hypothesis and heat conduction. Nernst, whose work on thermodynamics Hilbert had been following with interest,[190] spoke about the kinetic theory of rigid bodies. Von Smoluchowski came from Krakow and lectured on the limits of validity of the second law of thermodynamics, a topic he had already addressed at the Münster meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte*. Sommerfeld came from Munich to talk about problems of free trajectories. Lorentz was invited from Leiden; he spoke on the applications of kinetic theory to the study of the motion of the electron. Einstein was also invited, but he could not attend.[191] Evidently this was for Hilbert a major event and he took pains to announce it very prominently on the pages of the *Physikalische Zeitschrift*, including rather lengthy and detailed abstracts of the expected lectures for the convenience of those who intended to attend.[192] After the meeting Hilbert also wrote a detailed report on the lectures in the *Jahresbericht der Deutschen Mathematiker-Vereiningung*[193]as well as the introduction to the published collection (Planck et al. 1914). Hilbert expressed the hope that the meeting would stimulate further interest, especially among mathematicians, and lead to additional involvement with the exciting world of ideas created by the new physics of matter.

That semester Hilbert also taught two courses on physical issues, one on the theory of the electron and another on the principles of mathematics, quite similar to his 1905 course on the axiomatic method and including a long section on the axiomatization of physics as well. Hilbert's lectures on electron theory emphasized throughout the importance of the Lorentz transformations and of Lorentz covariance, and continually referred back to the works of Minkowski and Born. Hilbert stressed the need to

189 *Jahresbericht der Deutschen Mathematikervereinigung* 22 (1913), 27.

190 In January 1913, Hilbert had lectured on Nernst's law of heat at the Göttingen Physical Society (*Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 590). See also a remark added in Hilbert's handwriting in (Hilbert 1905a, 167).

191 Cf. Einstein to Hilbert, 4 October 1912 (CPAE 5, Doc. 417).

192 *Physikalische Zeitschrift* 14 (1913), 258–264. Cf. also (Born 1913).

193 *Jahresbericht der Deutschen Mathematiker-Vereinigung* 22 (1913), 53–68, which includes abstract of all the lectures. Cf. also *Jahresbericht der Deutschen Mathematiker-Vereinigung* 23 (1914), 41.

formulate unified theories in physics, and to explain all physical processes in terms of motion of points in space and time.[194] From this reductionistic point of view, the theory of the electron would appear as the most appropriate foundation of all of physics.[195] However, given the difficulty of explicitly describing the motion of, and the interactions among, several electrons, Hilbert indicated that the model provided by kinetic theory had to be brought to bear here. He thus underscored the formal similarities between mechanics, electrodynamics and the kinetic theory of gases. In order to describe the conduction of electricity in metals, he developed a mechanical picture derived from the theory of gases, which he then later wanted to substitute by an electrodynamical one.[196] Hilbert stressed the methodological motivation behind his quest after a unified view of nature, and the centrality of the demand for universal validity of the Lorentz covariance, in the following words:

> But if the relativity principle [i.e., invariance under Lorentz transformations] is valid, then it is so not only for electrodynamics, but for the whole of physics. We would like to consider the possibility of reconstructing the whole of physics in terms of as few basic concepts as possible. The most important concepts are the concept of force and of rigidity. From this point of view the electrodynamics would appear as the foundation of all of physics. But the attempt to develop this idea systematically must be postponed for a later opportunity. In fact, it has to start from the motion of one, of two, etc. electrons, and there are serious difficulties on the way to such an undertaking. The corresponding problem for Newtonian physics is still unsolved for more than two bodies.[197]

When looking at the kind of issues raised by Hilbert in this course, one can hardly be surprised to discover that somewhat later Gustav Mie's theory of matter eventually attracted his attention. Thus, for instance, Hilbert explained that in the existing theory of electrical conductivity in metals, only the conduction of electricity—which itself depends on the motion of electrons—has been considered, while assuming that the electron satisfies both Newton's second law, $F = ma$, and the law of collision as a perfectly elastic spherical body (as in the theory of gases).[198] This approach assumes that the magnetic and electric interactions among electrons are described correctly enough in these mechanical terms as a first approximation.[199] However, if we wish to investigate with greater exactitude the motion of the electron, while at the same time

---

194 "Alle physikalischen Vorgänge, die wir einer axiomatischen Behandlung zugängig machen wollen, suchen wir auf Bewegungsvorgänge an Punktsystem in Zeit und Raum zu reduzieren." (Hilbert 1913b, 1)

195 "Die Elektronentheorie würde daher von diesem Gesichtpunkt aus das Fundament der gesamten Physik sein." (Hilbert 1913b, 13)

196 "Unser nächstes Ziel ist, eine Erklärung der Elektrizitätsleitung in Metallen zu gewinnen. Zu diesem Zwecke machen wir uns von der Elektronen zunächst folgendes *der Gastheorie entnommene mechanische Bild*, das wir später durch ein elektrodynamisches ersetzen werden." (Hilbert 1913b, 14)

197 "Die wichtigsten Begriffe sind die der *Kraft* und der *Starrhei*t. Die Elektronentheorie würde daher von diesem Gesichtspunkt aus das Fundament der gesamten Physik sein. Den Versuch ihres systematischen Aufbaues verschieben wir jedoch auf später; er hätte von der Bewegung eines, zweier Elektronen u.s.w. auszugehen, und ihm stellen sich bedeutende Schwierigkeiten in der Weg, da schon die entsprechenden Probleme der Newtonschen Mechanik für mehr als zwei Körper ungelöst sind." (Hilbert 1913a, 1913c)

preserving the basic conception of the kinetic theory based on colliding spheres, then we should also take into account the field surrounding the electron and the radiation that is produced with each collision. We are thus led to investigate the influence of the motion of the electron on the distribution of energy in the free aether, or in other words, to study the theory of radiation from the point of view of the mechanism of the motion of the electron. In his 1912 lectures on the theory of radiation, Hilbert had already considered this issue, but only from a "phenomenological" point of view. This time he referred to Lorentz's work as the most relevant one.[200] From Lorentz's theory, he said, we can obtain the electrical force induced on the aether by an electron moving on the x-axis of a given coordinate system.

Later on, Hilbert returned once again to the mathematical difficulties implied by the basic assumptions of the kinetic model. When speaking of clouds of electrons, he said, one assumes the axioms of the theory of gases and of the theory of radiation. The *n*-electron problem, he said, is even more difficult than that of the *n*-bodies, and in any case, we can only speak here of averages. Hilbert thus found it more convenient to open his course by describing the motion of a single electron, and, only later on, to deal with the problem of two electrons.

In discussing the behavior of the single electron, Hilbert referred to the possibility of an electromagnetic reduction of all physical phenomena, freely associating ideas developed earlier in works by Mie and by Max Abraham. The Maxwell equations and the concept of energy, Hilbert said, do not suffice to provide a foundation of electrodynamics; the concept of rigidity has to be added to them. Electricity has to be attached to a steady scaffold, and this scaffold is what we denote as an electron. The electron, he explained to his students, embodies the concept of a rigid connection of Hertz's mechanics. In principle at least it should be possible to derive all the forces of physics, and in particular the molecular forces, from these three ideas: Maxwell's equations, the concept of energy, and rigidity. However, he stressed, one phenomenon has so far evaded every attempt at an electrodynamic explanation: the phenomenon of

198 "In der bisherigen Theoire der Elektricitätsleitung in Metallen haben wir nur den Elektrizitätstransport, der durch die Bewegung der Elektronen selbst bedingt wird, in Betracht gezogen; unter der Annahme, dass d*ie Elektronen erstens dem Kraftgesetz* $K = mb$ *gehorchen und zweitens dasselbe Stossgesetz wie vollkommen harten elastischen Kugeln befolgen* (wie in der Gastheorie)." (Hilbert 1913b, 14)

199 "Auf die elektrischen und magnetische Wirkung der Elektronen aufeinander und auf die Atome sind wir dabei nicht genauer eingegangen, vielmehr haben wir angenommen, dass die gegenseitige Beeinflussung durch das Stossgesetz in erster Annäherung hinreichend genau dargestellt würde." (Hilbert 1913b, 14)

200 "Wollte man die Wirkung der Elektronenbewegung genauer verfolgen—jedoch immer noch unter Beibehaltung des der Gastheorie entlehnten Bildes stossender Kugeln—so müsste man *das umgebende Feld der Elektronen und die Strahlung* in Rechnung setzen, die sie bei jedem Zusammenstoß aussenden. Man wird daher naturgemäß darauf geführt, den Einfluss der Elektronenbewegung auf die Energieverteilung im freien aether zu untersuchen. Ich gehe daher dazu über, die *Strahlungstheorie,* die wir früher vom phänomenologischen Standpunkt aus kennen gelernt haben, aus dem Mechanismus der Elektronenbewegung verständlich zu machen. Eine diesbezügliche Theorie hat *H. A. Lorentz* aufgestellt." (Hilbert 1913b, 14)

gravitation.[201] Still, in spite of the mathematical and physical difficulties that he considered to be associated with a conception of nature based on the model underlying kinetic theory, Hilbert did not fully abandon at this stage the mechanistic approach as a possible one, and in fact he asserted that the latter is a necessary consequence of the principle of relativity.[202]

11.3 Axiomatization of Physics: 1913

In 1913 Hilbert gave a course very similar to the one taught back in 1905, and bearing the same name: "Elements and Principles of Mathematics."[203] The opening page of the manuscript mentions three main parts that the lectures intended to cover:

A. Axiomatic Method.
B. The Problem of the Quadrature of the Circle.
C. Mathematical Logic.

In the actual manuscript, however, one finds only two pages dealing with the problem of the quadrature of the circle. Hilbert explained that, for lack of time, this section would be omitted in the course. Only a short sketch appears, indicating the stages involved in dealing with the problem. The third part of the course, "*Das mathematisch Denken und die Logik*," discussed various issues such as the paradoxes of set theory, false and deceptive reasoning, propositional calculus (*Logikkalkül*), the concept of number and its axioms, and impossibility proofs. The details of the contents of this last part, though interesting, are beyond our present concern here. In the first part Hilbert discussed in detail, like in 1905, the axiomatization of several physical theories.

Like in 1905, Hilbert divided his discussion of the axiomatic method into three parts: the axioms of algebra, the axioms of geometry, and the axioms of physics. In his first lecture Hilbert repeated the definition of the axiomatic method:

---

201 "Auf die Maxwellschen Gleichungen und den Energiebegriff allein kann man die Elektrodynamik nicht gründen. Es muss noch der Begriff der *Starrhei*t hinzukommen; die Elektrizität muss an ein festes Gerüst angeheftet sein. Dies Gerüst bezeichnen wir als Elektron. In ihm ist der Begriff der starrer Verbindung der Hertzschen Mechanik verwirklicht. Aus den Maxwellschen Gleichungen, dem Energiebegriff und dem Starrheitsbegriff lassen sich, im Prinzip wenigstens, die vollständigen Sätze der Mechanik entnehmen, auf sie lassen sich die gesamten Kräfte der Physik, im Besonderen die Molekularkräfte zurückzuführen. Nur die Gravitation hat sich bisher dem Versuch einer elektrodynamischen Erklärung widersetzt." (Hilbert 1913b, 61–62)

202 "Es sind somit die zum Aufbau der Physik unentbehrlichen starren Körper nur in den kleinsten Teilen möglich; man könnte sagen: das Relativitätsprinzip ergibt also als notwendige Folge die Atomistik." (Hilbert 1913b, 65)

203 The lecture notes of this course, (Hilbert 1913c), are not found in the Göttingen collections. Peter Damerow kindly allowed me to consult the copy of the handwritten notes in his possession. The notes do not specify who wrote them. In *Nachlass* David Hilbert, (Cod. Ms. D. Hilbert, 520, 5), Hilbert wrote that notes of the course were taken by Bernhard Baule.

> The axiomatic method consists in choosing a domain and putting certain facts on top of it; the proof of these facts does not occupy us anymore. The classical example of this is provided by geometry.[204]

Hilbert also repeated the major questions that should be addressed when studying a given system of axioms for a determined domain: Are the axioms consistent? Are they mutually independent? Are they complete?[205] The axiomatic method, Hilbert declared, is not a new one; rather it is deeply ingrained in the human way of thinking.[206]

Hilbert's treatment of the axioms of physical theories repeats much of what he presented in 1905 (the axioms of mechanics, the principle of conservation of energy, thermodynamics, calculus of probabilities, and psychophysics), but at the same time it contains some new sections: one on the axioms of radiation theory, containing Hilbert's recently published ideas on this domain, and one on space and time, containing an exposition of relativity. I comment first on one point of special interest appearing in the section on mechanics.

In his 1905 course Hilbert had considered the possibility of introducing alternative systems of mechanics defined by alternative sets of axioms. As already said, one of the intended aims of Hilbert's axiomatic analysis of physical theories was to allow for changes in the existing body of certain theories in the eventuality of new empirical discoveries that contradict the former. But if back in 1905, Hilbert saw the possibility of alternative systems of mechanics more as a mathematical exercise than as a physically interesting task, obviously the situation was considerably different in 1913. This time Hilbert seriously discussed this possibility in the framework of his presentation of the axioms of Newtonian mechanics. As in geometry, Hilbert said, one could imagine for mechanics a set of premises different from the usual ones and, from a logical point of view, one could think of developing a "non-Newtonian Mechanics."[207] More specifically, he used this point of view to stress the similarities between mechanics and electrodynamics. He had already done something similar in 1905, but now his remarks had a much more immediate significance. I quote them here in some extent:

> One can now drop or partially modify particular axioms; one would then be practicing a non-Newtonian, non-Galileian, or non-Lagrangian mechanics.

---

204 "Die axiomatische Methode besteht darin, daß man ein Gebiet herausgreift und bestimmte Tatsachen an die Spitze stellt u. nun den Beweis dieser Tatsachen sich nicht weiter besorgt. Das Musterbeispiel hierfür ist die Geometrie." (Hilbert 1913c, 1)

205 Again, Hilbert is not referring here to the model-theoretical notion of completeness. See § 2.1.

206 "Die axiomatische Methode ist nicht neu, sondern in der menschlichen Denkweise tief begründet." (Hilbert 1913c, 5)

207 "Logisch wäre es natürlich auch möglich andere Def. zu Grunde zu liegen und so eine 'Nicht-Newtonsche Mechanik' zu begründen." An elaborate formulation of a non-Newtonian mechanics had been advanced in 1909 by Gilbert N. Lewis (1875–1946) and Richard C. Tolman (1881–1948), in the framework of an attempt to develop relativistic mechanics independently of electromagnetic theory (Lewis and Tolman 1909). Hilbert did not give here a direct reference to that work but it is likely that he was aware of it, perhaps through the mediation of one of his younger colleagues. (Hilbert 1913c, 91)

> This has a very special significance: electrodynamics has compelled us to adopt the view that our mechanics is only a limiting-case of a more general one. Should anyone in the past have thought by chance of defining the kinetic energy as:
>
> $$T = \mu \frac{1 - v^2}{v} \log \frac{1 + v}{1 - v},$$
>
> he would have then obtained the [equation of] motion of the electron, where μ is constant and depends on the electron's mass. If one ascribes to all of them [i.e., to the electrons] kinetic energy, then one obtains the theory of the electron, i.e., an essential part of electrodynamics. One can then formulate the Newtonian formula:
>
> $$ma = F$$
>
> But now the mass depends essentially on the velocity and it is therefore no more a physical constant. In the limit case, when the velocity is very small, we return to the classical physics....
>
> Lagrange's equations show how a point moves when the conditions and the forces are known. How these forces are created and what is their nature, however, this is a question which is not addressed.
>
> *Boltzmann* attempted to build the whole of physics starting from the forces; he investigated them, and formulated axioms. His idea was to reduce everything to the mere existence of central forces of repulsion or of attraction. According to Boltzmann there are only mass-points, mutually acting on each other, either attracting or repelling, over the straight line connecting them. Hertz was of precisely of the opposite opinion. For him there exist no forces at all; rigid bonds exist among the individual mass-points. Neither of these two conceptions has taken root, and this is for the simple reason that electrodynamics dominates all.
>
> The foundations of mechanics, and especially its goal, are not yet well established. Therefore it has no definitive value to construct and develop these foundations in detail, as has been done for the foundations of geometry. Nevertheless, this kind of foundational research has its value, if only because it is mathematically very interesting and of an inestimably high value.[208]

This passage illuminates Hilbert's conceptions by 1913. At the basis of his approach to physics stands, as always, the axiomatic method as the most appropriate way to examine the logical structure of a theory and to decide what are the individual assumptions from which all the main laws of the theory can be deduced. This deduction, however, as in the case of Lagrange's equation, is independent of questions concerning the ultimate nature of physical phenomena. Hilbert mentions again the mechanistic approach promoted by Hertz and Boltzmann, yet he admits explicitly, perhaps for the first time, that it is electromagnetism that pervades all physical phenomena. Finally, the introduction of Lagrangian functions from which laws of motion may be derived that are more general than the usual ones of classical mechanics was an idea that in the past might have been considered only as a pure mathematical exercise; now—Hilbert cared to stress—it has become a central issue in mechanics, given the latest advances in electrodynamics.

The last section of Hilbert's discussion of the axiomatization of physics addressed the issue of space and time, and in fact it was a discussion of the principle of relativity.[209] What Hilbert did in this section provides the most detailed evidence of his conceptions concerning the principle of relativity, mechanics and electrodynamics before

his 1915 paper on the foundations of physics. His presentation did not really incorporate any major innovations, yet Hilbert attempted to make the "new mechanics" appear as organically integrated into the general picture of physics that he was so eager to put forward at every occasion, and in which all physical theories appear as in principle axiomatized (or at least axiomatizable). Back in 1905, Hilbert had suggested, among the possible ways to axiomatize classical dynamics, defining space axiomatically by means of the already established axioms of geometry, and then expanding this definition with some additional axioms that define time. He suggested that something similar should be done now for the new conception of space and time, but that the axioms defining time would clearly have to change. He thus assumed the axioms of Euclidean geometry and proceeded to redefine the concept of time using a "light pendulum." Hilbert then connected the axiomatically constructed theory with the additional empirical consideration it was meant to account for, namely, the outcome of the Michelson-Morley experiment when the values $\vartheta = 0, \pi/2, \pi,$ are measured in the formula describing the velocity of the ray-light γ_ϑ in the pendulum:

---

208 "Man kann nun gewisse Teile der Axiome fallen lassen oder modifizieren; dann würde man also "*Nicht-Newtonsche*," od. "*Nicht-Galileische*", od. "*Nicht-Lagrangesche*" *Mechanik* treiben.
Das hat ganz besondere Bedeutung: Durch die Elektrodynamik sind wir zu der Auffassung gezwungen worden, daß unsere Mechanik nur eine Grenzfall einer viel allgemeineren Mechanik ist. Wäre jemand früher zufällig darauf gekommen die kinetisch Energie zu definieren als:

$$T = \mu \frac{1 - v^2}{v} \log \frac{1 + v}{1 - v},$$

so hatte er die Bewegung eines Elektrons, wo μ eine Constante der elektr. Masse ist. Spricht man ihnen allen kinetisch Energie zu, dann hat man die Elektronentheorie d.h. einen wesentlichen Teil der Elektrodynamik. Dann kann man die Newtonschen Gleichungen aufstellen:

$$mb = K$$

Nun hängt aber die Masse ganz wesentlich von der Geschwindigkeit ab und ist keine physikalische Constante mehr. Im Grenzfall, daß die Geschwindigkeit sehr klein ist, kommt man zu der alten Mechanik zurück. (Cf. H. Stark "Experimentelle Elektrizitätslehre," S. 630).
Die Lagrangesche Gleichungen geben die Antwort wie sich ein Punkt bewegt, wenn man die Bedingungen kennt und die Kräfte. Wie diese Kräfte aber beschaffen sind und auf die Natur die Kräfte selbst gehen sie nicht ein.
Boltzmann hat versucht die Physik aufzubauen indem er von der Kräften ausging; er untersuchte diese, stellte Axiome auf u. seine Idee war, alles auf das bloße Vorhandensein von Kräften, die zentral abstoßend oder anziehend wirken sollten, zurückzuführen. Nach Boltzman gibt es nur Massenpunkte die zentral gradlinig auf einander anzieh. od. abstoßend wirkend.
Hertz hat gerade den entgegengesetzten Standpunkt. Für ihn gibt es überhaupt keine Kräfte; starre Verbindungen sind zwischen den einzelnen Massenpunkten.
Beide Auffassungen haben sich nicht eingebürgert, schon aus dem einfachen Grunde, weil die Elektrodynamik alles beherrscht.
Die Grundlagen der Mechanik und besonders die Ziele stehen noch nicht fest, so daß es auch noch nicht definitiven Wert hat die Grundlagen in den einzelnen Details so auf- und ausbauen wie die *Grundlagen der Geometrie*. Dennoch behalten die axiomatischen Untersuchungen ihren Wert, schon deshalb, weil sie mathematisch sehr interessant und von unschätzbar hohen Werte sind." (Hilbert 1913c, 105–108)

$$\gamma_\vartheta = \sqrt{\frac{\xi^2+\eta^2}{t^2}} = \sqrt{\cos^2\vartheta + \sin^2\vartheta - 2v\cos\vartheta + v^2} = \left[1 - 2v\cos\vartheta + v^2\right]^{\frac{1}{2}}.$$

Hilbert stressed the similarities between the situation in this case, and in the case in geometry, when one invokes Gauss's measurement of angles in the mountain triangle for determining the validity of Euclidean geometry in reality. In his earlier lectures, Hilbert had repeatedly mentioned this experiment as embodying the empirical side of geometry. The early development of relativity theory had brought about a deep change in the conception of time, but Hilbert of course could not imagine that the really significant change was still ahead. To the empirical discovery that triggered the reformulation of the concept of time, Hilbert opposed the unchanged conception of space instantiated in Gauss's experiment. He thus said:

> Michelson set out to test the correctness of these relations, which were obtained working within the old conception of time and space. The [outcome of his] great experiment is that these formulas do not work, whereas Gauss had experimentally confirmed (i.e., by measuring the Hoher Hagen, the Brocken, and the Inselsberg) that in Euclidean geometry, the sum of the angles of a triangle equals two right angles.[210]

Although he spoke here of an old conception of space and time, Hilbert was referring to a change that actually affected only time. From the negative result of Michelson's experiment, one could conclude that the assumption implied by the old conception—according to which, the velocity of light measured in a moving system has different values in different directions—leads to contradiction. The opposite assumption was thus adopted, namely that the velocity of light behaves with respect to moving systems as it had been already postulated for stationary ones. Hilbert expresses this as a further axiom:

---

209 The following bibliographical list appears in the first page of this section (Hilbert 1913c, 119):
M. Laue Das Relativitätsprinzip 205 S.
M. Planck 8 Vorlesungen über theoretische Physik 8. Vorlesung S. 110–127
A. Brill Das Relativitätsprinzip: eine Einführung in die Theorie 28 S.
H. Minkowski Raum und Zeit XIV Seiten
Beyond this list, together with the manuscript of the course, in the same binding, we find some additions, namely, (1) a manuscript version of Minkowski's famous work (83 pages in the same handwriting as the course itself), (2) the usual preface of A. Gutzmer, appearing as an appendix, and (3) two pages containing a passage copied form Planck's *Vorlesungen*.

210 "Diese aus der alten Auffassung von Raum und Zeit entspringende Beziehung hat Michelson auf ihre Richtigkeit geprüft. Das große Experiment ist nun das, daß diese Formel nicht stimmt, während bei der Euklidischen Geometrie Gauss durch die bestimmte Messung Hoher Hagen, Brocken, Inselsberg bestätigte, daß die Winkelsumme im Dreieck 2 Rechte ist."
On p. 128 Hilbert explained the details of Michelson's calculations, namely, the comparison of velocities at different angles via the formula:

$$\frac{1}{\gamma_\vartheta} + \frac{1}{\gamma_{\vartheta+\pi}} = \left(1 - 2v\cos\vartheta + \vartheta^2\right)^{-\frac{1}{2}} + \left(1 + 2v\cos\vartheta + \vartheta^2\right)^{-\frac{1}{2}} = 2 + v^2\left(3\cos^2\vartheta - 1\right) + \ldots$$

where the remaining terms are of higher orders. (Hilbert 1913c, 124)

> Also in a moving system, the velocity of light is identical in all directions, and in fact, identical to that in a stationary system. The moving system has no priority over the first one.[211]

Now the question naturally arises: what is then the true relation between time as measured in the stationary system and in the moving one, t and τ respectively? Hilbert answered this question by introducing the Lorentz transformations, which he discussed in some detail, including the limiting properties of the velocity of light,[212] and the relations with a third system, moving with yet a different uniform velocity.

11.4 Electromagnetic Oscillations: 1913–1914

In the winter semester of 1913–1914, Hilbert lectured on electromagnetic oscillations. As he had done many times in the past, Hilbert opened by referring to the example of geometry as a model of an experimental science that has been transformed into a purely mathematical, and therefore a "theoretical science," thanks to our thorough knowledge of it. Foreshadowing the wording he would use later in his axiomatic formulation of the general theory of relativity, Hilbert said:

> From antiquity the discipline of geometry is a part of mathematics. The experimental grounds necessary to build it are so suggestive and generally acknowledged, that from the outset it has immediately appeared as a theoretical science. I believe that the highest glory that such a science can attain is to be assimilated by mathematics, and that theoretical physics is presently on the verge of attaining this glory. This is valid, in the first place for the relativistic mechanics, or four-dimensional electrodynamics, which belong to mathematics, as I have been already convinced for a long time.[213]

Hilbert's intensive involvement with various physical disciplines over the last years had only helped to strengthen an empirical approach to geometry rather than promoting some kind of formalist views. But as for his conceptions about physics itself, by the end of 1913 his new understanding of the foundational role of electrodynamics was becoming only more strongly established in his mind, at the expense of his old mechanistic conceptions. The manuscript of this course contains the first doc-

---

211 "Es zeigt sich also, daß unsere Folgerung der alten Auffassung, daß die Lichtgeschwindigkeit im bewegtem System nach verschiedenen Richtungen verschieden ist, auf Widerspruch führt. Wir nehmen deshalb an: *Auch im bewegtem System ist die Lichtgeschwindigkeit nach allem Seiten gleich groß, und zwar gleich der im ruhenden. Das bewegte System hat vor dem alten nicht voraus*." (Hilbert 1913c, 128–129)

212 "Eine größen Geschwindigkeit als die Lichtgeschwindigkeit kann nicht vorkommen." (Hilbert 1913c, 132)

213 "Seit Alters her ist die Geometrie eine Teildisziplin der Mathematik; die experimentelle Grundlagen, die sie benutzen muss, sind so naheliegend und allgemein anerkannt, dass sie von vornherein und unmittelbar als theoretische Wissenschaft auftrat. Nun glaube ich aber, dass es der höchste Ruhm einer jeden Wissenschaft ist, von der Mathematik assimiliert zu werden, und dass auch die theoretische Physik jetzt im Begriff steht, sich diesen Ruhm zu erwerben. In erster Linie gilt dies von der Relativitätsmechanik oder vierdimensionalen Elektrodynamik, von deren Zuhörigkeit zur Mathematik ich seit langem überzeugt bin." (Hilbert 1913–1914, 1)

umented instance where Hilbert seems to allude to Mie's ideas and, indeed, it is among the earliest explicit instances of a more decided adoption of electrodynamics, rather than mechanics, as the possible foundation for all physical theories. At the same time, the whole picture of mathematics was becoming ever more hierarchical and organized into an organic, comprehensive edifice, of which theoretical physics is also an essential part. Hilbert thus stated:

> In the meantime it looks as if, finally, theoretical physics completely arises from electrodynamics, to the extent that every individual question must be solved, in the last instance, by appealing to electrodynamics. According to what method each mathematical discipline more predominantly uses, one could divide mathematics (concerning contents rather than form) into one-dimensional mathematics, i.e., arithmetic; then function theory, which essentially limits itself to two dimensions; then geometry, and finally four-dimensional mechanics.[214]

In the course itself, however, Hilbert did not actually address in any concrete way the kind of electromagnetic reduction suggested in its introduction, but rather, it continued, to a certain extent, his previous course on electron theory. In the first part Hilbert dealt with the theory of dispersion of electrons, seen as a means to address the *n*-electron problem. Hilbert explained that the role of this problem in the theory of relativity is similar to that of the *n*-body problem in mechanics. In the previous course he had shown that the search for the equations of motion for a system of electrons leads to a very complicated system of integro-differential equations. A possibly fruitful way to address this complicated problem would be to integrate a certain simplified version of these equations and then work on generalizing the solutions thus obtained. In classical mechanics the parallel simplification of the *n*-body problem is embodied in the theory of small oscillations, based on the idea that bodies cannot really attain a state of complete rest. This idea offers a good example of a possible way forward in electrodynamics, and Hilbert explained that, indeed, the elementary theory of dispersion was meant as the implementation of that idea in this field. Thus, this first part of the course would deal with it.[215]

---

214 "Es scheint indessen, als ob die theoretische Physik schliesslich ganz und gar in der Elektrodynamik aufgeht, insofern jede einzelne noch so spezielle Frage in letzter Instanz an die Elektrodynamik appellieren muss. Nach den Methoden, die die einzelnen mathematischen Disziplinen vorwiegend benutzen, könnte man alsdann – mehr inthaltlich als formell – die Mathematik einteilen in die eindimensionale Mathematik, die Arithmetik, ferner in die Funktionentheorie, die sich im wesentlichen auf zwei Dimensionen beschränkt, in die Geometrie, und schliesslich in die vierdimensionale Mechanik." (Hilbert 1913–1914, 1)

215 "So wenig man schon mit dem *n*-Körperproblem arbeiten kann, so wäre es noch fruchtloser, auf die Behandlung des *n*-Elektronenproblemes einzugehen. Es handelt sich vielmehr für uns darum, das *n*-Elektronenproblem zu verstümmeln, die vereinfachte Gleichungen zu integrieren und von ihren Lösungen durch Korrekturen zu allgemeineren Lösungen aufzusteigen. Die gewöhnliche Mechanik liefert uns hierfür ein ausgezeichnetes Vorbild in der Theorie der kleinen Schwingungen; die Vereinfachung des *n*-Körperproblems besteht dabei darin, dass die Körper sich nur wenig aus festen Ruhelagen entfernen dürfen. In der Elektrodynamik gibt es ein entsprechendes Problem, und zwar würde ich die Theorie der Dispersion als das dem Problem der kleinen Schwingungen analoge Problem ansprechen." (Hilbert 1913–1914, 2)

In the second part of the course Hilbert dealt with the magnetized electron. He did not fail to notice the difficulties currently affecting his reductionist program. At the same time he stressed the value of an axiomatic way of thinking in dealing with such difficulties. He thus said:

> We are really still very distant from a full realization of our leading idea of reducing all physical phenomena to the *n*-electron problem. Instead of a mathematical foundation based on the equations of motion of the electrons, we still need to adopt partly arbitrary assumptions, partly temporary hypothesis, that perhaps one day in the future might be confirmed. We also must adopt, however, certain very fundamental assumptions that we later need to modify. This inconvenience will remain insurmountable for a long time. What sets our presentation apart from that of others, however, is the insistence in making truly explicit all its assumptions and never mixing the latter with the conclusions that follow from them.[216]

Hilbert did not specify what assumptions he meant to include under each of the three kinds mentioned above.Yet it would seem quite plausible to infer that the "very fundamental assumptions," that must be later modified, referred in some way or another to physical, rather than purely mathematical, assumptions, and more specifically, to the atomistic hypothesis, on which much of his own physical conceptions had hitherto been based. An axiomatic analysis of the kind he deemed necessary for physical theories could indeed compel him to modify even his most fundamental assumptions if necessary. The leading principle should remain, in any case, to separate as clearly as possible the assumptions of any particular theory from the theorems that can be derived in it. Thus, the above quotation suggests that if by this time Hilbert had not yet decided to abandon his commitment to the mechanistic reductionism and its concomitant atomistic view, he was certainly preparing the way for that possibility, should the axiomatic analysis convince him of its necessity.

In the subsequent lectures in this course, Hilbert referred more clearly to ideas of the kind developed in Mie's theory, without however explicitly mentioning his name (at least according to the record of the manuscript). Outside ponderable bodies, which are composed of molecules, Hilbert explained, the Maxwell equations are valid. He formulated them as follows:

$$\mathrm{curl}M - \frac{\partial e}{\partial t} = \rho\varpi;\ \mathrm{div}e = \rho$$
$$\mathrm{curl}e + \frac{\partial M}{\partial t} = 0;\ \mathrm{div}\ M = 0$$

216 "Von der Verwirklichung unseres leitenden Gedankens, alle physikalischen Vorgänge auf das *n*-Elektronenproblem zurückzuführen, sind wir freilich noch sehr weit entfernt. An Stelle einer mathematischen Begründung aus den Bewegungsgleichungen der Elektronen müssen vielmehr noch teils willkürliche Annahmen treten, teils vorläufige Hypothesen, die später einmal begründet werden dürften, teils aber auch Annahmen ganz prinzipieller Natur, die sicher später modifiziert werden müssen. Dieser Übelstand wird noch auf lange Zeit hinaus unvermeidlich sein. Unsere Darstellung soll sich aber gerade dadurch auszeichnen, dass die wirklich nötigen Annahmen alle ausdrücklich aufgeführt und nicht mit ihren Folgerungen vermischt werden." (Hilbert 1913–1914, 87–88)

This is also how the equations are formulated in Born's article of 1910, the text on which Hilbert was basing this presentation. But Hilbert asserted here for the first time that the equations are valid also inside the body. And he added:

> Inside the body, however, the vectors e and M are very different, since the energy density is always different from zero inside the sphere of the electron, and these spheres undergo swift oscillations. It would not help us to know the exact value of the vector fields inside the bodies, since we can only observe mean values.[217]

Hilbert thus simply stated that the Maxwell equations inside the body should be rewritten as:

$$\mathrm{curl}\,\overline{M}\left(-\frac{\partial \bar{e}}{\partial t}\right) = \bar{\rho}\,\overline{\varpi};\ \mathrm{div}\, e = \rho$$

$$\mathrm{curl}\,\bar{e} + \frac{\partial \overline{M}}{\partial t} = 0;\ \mathrm{div}\ M = 0$$

where overstrike variables indicate an average value over a space region.

Hilbert went on to discuss separately and in detail specific properties of the conduction-, polarization- and magnetization-electrons. He mentioned Lorentz as the source for the assumption that these three kinds of electrons exist. This assumption, he said, is an "assumption of principle" that should rather be substituted by a less arbitrary one.[218] By saying this, he was thus not only abiding by his self-imposed rules that every particular assumption must be explicitly formulated, but he was also implicitly stressing once again that physical assumptions about the structure of matter are of a different kind than merely mathematical axioms, that they should be avoided whenever possible, and that they should eventually be suppressed altogether.

In a later section of his lecture, dealing with diffuse radiation and molecular forces, Hilbert addressed the problem of gravitation from an interesting point of view that, once again, would seem to allude to the themes discussed by Mie, without however explicitly mentioning his name. Hilbert explained that the problem that had originally motivated the consideration of what he called "diffuse electron oscillations" (a term he did not explain) was the attempt to account for gravitation. In fact, he added, it would be highly desirable—from the point of view pursued in the course—to explain gravitation based on the assumption of the electromagnetic field and the Maxwell equations, together with some auxiliary hypotheses, such as the existence of rigid bodies. The idea of explaining gravitation in terms of "diffuse radiation of a given wavelength" was, according to Hilbert, closely related to an older idea first raised by

---

217 "Diese Gleichungen gelten sowohl innerhalb wie *ausserhalb des Körpers*. Im innern des Körpers werden aber die Vektoren E und M sich räumlich und zeitlich sehr stark ändern, da die Dichte der Elektrizität immer nur innerhalb der Elektronenkugeln von Null verschieden ist und diese Kugeln rasche Schwingungen ausführen. Es würde uns auch nicht helfen, wenn wir innerhalb des Körpers die genauen Werte der Feldvektoren kennen würden; denn zur Beobachtung gelangen doch nur Mittelwerte." (Hilbert 1913–1914, 89)

218 "Wir machen nur eine reihe von Annahmen, die zu den prinzipiellen gehören und später wohl durch weniger willkürlich scheinende ersetzt werden können." (Hilbert 1913–1914, 90)

Georges-Louis Le Sage (1724–1803). The latter was based on the assumption that a great number of particles move in space with a very high speed, and that their impact with ponderable bodies produces the phenomenon of weight.[219] However, Hilbert explained, more recent research has shown that an explanation of gravitation along these lines is impossible.[220] Hilbert was referring to an article published by Lorentz in 1900, showing that no force of the form $1/r^2$ is created by "diffuse radiation" between two electrical charges, if the distance between them is large enough when compared to the wavelength of the radiation in question (Lorentz 1900).[221]

And yet in 1912, Erwin Madelung had readopted Lorentz's ideas in order to calculate the force produced by radiation over short distances and, eventually, to account for the molecular forces in terms of radiation phenomena (Madelung 1912). Madelung taught physics at that time in Göttingen and, as we saw, he had attended Hilbert's 1912 advanced seminar on kinetic theory. Hilbert considered that the mathematical results obtained by him were very interesting, even though their consequences could not be completely confirmed empirically. Starting from the Maxwell equations and some simple, additional hypotheses, Madelung determined the value of an attraction force that alternatively attains positive and negative values as a function of the distance.[222]

As a second application of diffuse radiation, Hilbert mentioned the possibility of deriving Planck's radiation formula without recourse to quantum theory. Such a derivation, he indicated, could be found in two recent articles of Einstein, one of them (1910) with Ludwig Hopf (1884–1939) and the second one (1913) with Otto Stern (1888–1969).

Hilbert's last two courses on physics, before he began developing his unified theory and became involved with general relativity, were taught in the summer semester

219 LeSage's corpuscular theory of gravitation, originally formulated in 1784, was reconsidered in the late nineteenth century by J.J. Thomson. On the Le Sage-Thomson theory see (North 1965, 38–40; Roseaveare 1982, 108–112). For more recent discussions, cf. also (Edwards 2002).

220 "Das Problem, das zunächst die Betrachtung diffuser Elektronenschwingungen anregte, war die Erklärung der Gravitation. In der Tat muss es ja nach unserem leitenden Gesichtspunkte höchst wünschenswert erscheinen, die Gravitation allein aus der Annahme eines elektromagnetischen Feldes sowie er Maxwellschen Gleichungen und gewisser einfacher Zusatzhypothesen, wie z.B. die Existenz starrer Körper eine ist, zu erklären. Der Gedanke, den Grund für die Erscheinung der Gravitation in einer diffusen Strahlung von gewisser Wellelänge zu suchen, ähnelt entfernt einer Theorie von Le Sage, nach der unzählige kleine Partikel sich mit grosser Geschwindigkeit im Raume bewegen sollen und durch ihren Anprall gegen die ponderablen Körper die Schwere hervorbringen. Wie in dieser theorie ein Druck durch bewegte Partikel auf die Körper ausgeübt wird, hat man jetzt den modernen Versuch unternommen, den Strahlungsdruck für die Erklärung der Gravitation dienstbar zu machen." (Hilbert 1913–1914, 107–108)

221 On this theory, see (McCormmach 1970, 476–477).

222 "Die mathematischen Ergebnisse dieser Arbeit sind von grossem Interesse, auch wenn sich die Folgerungen nicht sämtlich bewähren sollten. Es ergibt sich nämlich allein aus den Maxwellschen Gleichungen und einfachen Zusatzhypothese eine ganz bestimmte Attraktionskraft, die als Funktion der Entfernung periodisch positiv und negativ wird." (Hilbert 1913–1914, 108)

of 1914 (statistical mechanics) and the following winter semester, 1914–1915 (lectures on the structure of matter).[223]

12. BROADENING PHYSICAL HORIZONS - CONCLUDING REMARKS

The present chapter has described Hilbert's intense and wide-ranging involvement with physical issues between 1910 and 1914. His activities comprised both published work and courses and seminars. In the published works, particular stress was laid on considerably detailed axiomatic analysis of theories, together with the application of the techniques developed by Hilbert himself in the theory of linear integral equations. The courses and seminars, however, show very clearly that Hilbert was not just looking for visible venues in which to display the applicability of these mathematical tools. Rather, they render evident the breadth and depth of his understanding of, and interest in, the actual physical problems involved.

Understanding the mixture of these two components—the mathematical and the physical—helps us to understand how the passage from mechanical to electromagnetic reductionism was also the basis of Hilbert's overall approach to physics, and particularly of his fundamental interest in the question of the structure of matter. In spite of the technical possibilities offered by the theory of integral equations in the way to solving specific, open problems in particular theories, Hilbert continued to be concerned about the possible justification of introducing probabilistic methods in physical theories at large. If the phenomenological treatment of theories was only a preliminary stage on the way to a full understanding of physical processes, it turned out that also those treatments based on the atomistic hypothesis, even where they helped reach solutions to individual problems, raised serious foundational questions that required further investigation into the theory of matter as such. Such considerations were no doubt a main cause behind Hilbert's gradual abandonment of mechanical reductionism as a basic foundational assumption.

This background should suffice to show the extent to which his unified theory of 1915 and the concomitant incursion into general theory of relativity were organically connected to the life-long evolution of his scientific horizon, and were thus anything but isolated events. In addition to this background, there are two main domains of ideas that constitute the main pillars of Hilbert's theory and the immediate catalysts for its formulation. These are the electromagnetic theory of matter developed by Gustav Mie starting in 1912, on the one hand, and the efforts of Albert Einstein to generalize the principle of relativity, starting roughly at the same time.

223 The winter semester, 1914–1915 course is registered in the printed version of the *Verzeichnis der Vorlesungen an der Georg-August-Universität zu Göttingen* (1914–1915, on p. 17) but no notes seem to be extant.

REFERENCES

Abraham, Max. 1902 "Dynamik des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse* (1902), 20–41.

———. 1903 "Prinzipien der Dynamik des Elektrons." *Annalen der Physik* 10, 105–179.

Alexandrov, Pavel S. (ed.). 1979. *Die Hilbertsche Probleme*. (*Ostwalds Klassiker der exakten Wissenschaften*, vol. 252.) Leipzig: Akad. Verlagsgesellschaft. (German edition of the Russian original.)

Arabatzis, Theodore. 1996 "Rethinking the 'Discovery' of the Electron." *Studies in History and Philosophy of Modern Physics* 27, 405–435.

Barbour, Julian. 1989. *Absolute or Relative Motion. A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories*. Cambridge: Cambridge University Press.

Barkan, Diana. 1993 "The Witches' Sabbath: The First International Solvay Congress in Physics." *Science in Context* 6, 59–82.

Blum, Petra. 1994. *Die Bedeutung von Variationsprinzipien in der Physik für David Hilbert*. Unpublished State Examination. Mainz: Johannes Gutenberg-Universität.

Blumenthal, Otto. 1935 "Lebensgeschichte." (Hilbert 1932–1935, vol. 3, 387–429)

Bohlmann, Georg. 1900 "Ueber Versicherungsmathematik." In F. Klein and E. Riecke (eds.), *Über angewandte Mathematik und Physik in ihrer Bedeutung für den Unterricht an der höheren Schulen*. Leipzig: Teubner, 114–145.

Boltzmann, Ludwig. 1872. "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen." *Sitzungsberichte Akad. Wiss. Vienna* 66: 275–370. (*Wissenschaftliche Abhandlungen*, vol. 1, 316–402. English translation in (Brush 1966).

———. 1897. *Vorlesungen ueber die Principien der Mechanik*. Leipzig: Verlag von Ambrosius Barth. English translation of the 'Introduction' in (Boltzmann 1974, 223–254).

———. 1899 "Über die Entwicklung der Methoden der theoretischen Physik in neuerer Zeit." In Boltzmann *Populäre Schriften*. Leipzig: J.A. Barth (1905), 198–277). English translation in (Boltzmann 1974, 77–100.)

———. 1900 "Die Druckkräfte in der Hydrodynamik und die Hertzsche Mechanik." *Annalen der Physik* 1, 673–677. (*Wissenschaftliche Abhandlungen*, 3 vols. Leipzig (1909), vol. 3, 665–669. (Chelsea reprint, New York, 1968.)

———. 1974. *Theoretical Physics and Philosophical Problems. Selected Writings*. (Translated by Paul Foulkes, edited by Brian McGuiness, Foreword by S.R. de Groot.) Dordrecht: Reidel.

Born, Max. 1913. "Zur kinetische Theorie der Materie. Einführung zum Kongreß in Göttingen." *Die Naturwissenschaften* 1, 297–299.

———. 1922. "Hilbert und die Physik." *Die Naturwissenschaften* 10, 88–93. (Reprinted in Born 1963, Vol. 2, 584–598.)

———. 1963. *Ausgewählte Abhandlungen*. Göttingen: Vandenhoek & Ruprecht.

———. 1978. *My Life: Recollections of a Nobel Laureate*. New York: Scribner's.

Breitenberg, Ernst. 1984. "Gauss's Geodesy and the Axiom of Parallels." *Archive for History of Exact Sciences* 31, 273–289.

Browder, Felix E. 1976. *Mathematical Developments Arising from Hilbert Problems*. (*Symposia in Pure Mathematics*, vol. 28.) Providence: American Mathematical Society.

Brush, Stephen G. (ed.). 1966. *Kinetic Theory* Vol. 2, *Irreversible Processes*. Oxford: Pergamon Press, 88–175.

Brush, Stephen G. 1976. *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century*. Amsterdam/New York/Oxford: North Holland Publishing House.

Bucherer, Alfred Heinrich. 1903. *Elemente der Vektor-Analysis. Mit Beispielen aus der theroretischen Physik*. Leipzig: Teubner.

Corry, Leo. 2003. *Modern Algebra and the Rise of Mathematical Structures*. Basel/Boston: Birkhäuser. 2nd revised edition. (1st ed.: *Science Networks*, vol. 17, 1996)

———. 2004. *David Hilbert and the Axiomatization of Physics, 1898–1918: From "Grundlagen der Geometrie" to "Grundlagen der Physik"*. Dordrecht: Kluwer.

CPAE 2. 1989. John Stachel, David C. Cassidy, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 2. *The Swiss Years: Writings, 1900–1909*. Princeton: Princeton University Press.

CPAE 3. 1993. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 3. *The Swiss Years: Writings, 1909–1911*. Princeton: Princeton University Press.

CPAE 4. 1995. Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 5. 1993. Martin J. Klein, A. J. Kox, and Robert Schulmann (eds.), *The Collected Papers of Albert Einstein*. Vol. 5. *The Swiss Years: Correspondence, 1902–1914*. Princeton: Princeton University Press.

Crowe, Michael J. 1967. *A History of Vector Analysis. The Evolution of the Idea of a Vectorial System*. University of Notre Dame Press.

Darboux, Gaston. 1875. "Sur la composition des forces en statique." *Bull. Sci. Math. Astr.* 18: 281–288.

Darrigol, Olivier. 2000. *Electrodynamics from Ampère to Einstein*. Chicago: The University of Chicago Press.

Dorier, Jean Luc. 1995. "A General Outline of the Genesis of Vector Space Theory." *Historia Mathematica* 22: 227–261.

Du Bois-Reymond, Emil. 1872. "Ueber die Grenzen des Naturerkennens." Vortrag in der 2. öffentlichen Sitzung der 45. *Versammlung deutscher Naturforscher und Ärzte*, Leipzig, 14 August 1872.

Dugac, Pierre. 1976. *Richard Dedekind et les fondements des mathématiques*. Paris: Vrin.

Duhamel, Jean Marie Constant. 1853–1854. *Cours de mécanique de l'École polytechnique*, 2nd.ed. Paris: Mallet-Bachelier.

Edwards, Mathew R. (ed.). 2002. *Pushing Gravity. New Perspectives on Le Sage's Theory of Gravitation*. Montreal: Apeiron.

Ehrenfest, Paul. 1904. "Die Bewegung Starrer Körper in Flüssigkeiten und die Mechanik von Hertz." In M. Klein (ed.), *Paul Ehrenfest. Collected Scientific Papers*. Amsterdam: North Holland (1959), 1–75.

———. 1911. "Welche Züge der Lichtquantenhypothese spielen in der Theorie der Wärmestrahlung eine wesentliche Rolle?" *Annalen der Physik* 36, 91–118.

Einstein, Albert. 1902. "Kinetischen Theorie der Wärmegleichgewichts und des zweiten Haupsatzes der Thermodynamik." *Annalen der Physik* 9: 417–433, (CPAE 2, Doc. 3).

Einstein, Albert and Ludwig Hopf. 1910. "Statistische Untersuchungen der Bewegung eines Resonators in einem Strahlungsfeld." *Annalen der Physik* 33, 1105–1115, (CPAE 3, Doc. 8).

Einstein, Albert and Otto Stern. 1913. "Einige Argumente für die Annahme einer molekularen Agitation beim absoluten Nullpunkt." *Annalen der Physik* 40: 551–560, (CPAE 4, Doc. 11).

Ewald, William. (ed.). 1999. *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, 2 vols. Oxford: Clarendon Press.

Ferreirós Domínguez, José. 1999. *Labyrinth of Thought. A History of Set Theory and its Role in Modern Mathematics*. Boston: Birkhäuser. (*Science Networks* Vol. 23.)

Föppl, August. 1901. *Vorlesungen über technische Mechanik*, 2nd. ed. Leipzig: Teubner.

Frege, Gottlob. 1903. *Grundgesetze der Arithmetik*, vol. 2. Jena: Pohle.

Frei, Günther. (ed.). 1985. *Der Briefwechsel David Hilbert-Felix Klein (1996–1919)*. Göttingen: Vandenhoeck & Ruprecht.

Gabriel, Gottfried. et al. (eds.). 1980. *Gottlob Frege - Philosophical and Mathematical Correspondence*. Chicago: The University of Chicago Press. (Abridged from the German edition by Brian McGuiness and translated by Hans Kaal.)

Gans, Richard. 1905. *Einführung in die Vektoranalysis. Mit Anwendungen auf die mathematische Physik*. Leipzig: Teubner.

Gibbs, Josiah Willard. 1902. *Elementary Principles of Statistical Mechanics*. New York: Scribner.

Gleason, A. 1952. "Groups without Small Subgroups." *Annals of Mathematics* 56: 193–212.

Goldberg, Stanley. 1970. "The Abraham Theory of the Electron: The Symbiosis of Experiment and Theory." *Archive for History of Exact Sciences* 7: 7–25.

Grattan-Guinness, Ivor. 2000. "A Sideways Look at Hilbert's Twenty-three Problems of 1900." *Notices American Mathematical Society* 47 (7): 752–757.

———. 2001. *Notices American Mathematical Society*, 48 (2): 167.

Gray, Jeremy J. (ed.). 2000. *The Hilbert Challenge*. New York: Oxford University Press.

Hamel, Georg. 1905. "Über die Zusammensetzung von Vektoren." *Zeitschrift für Mathematik und Physik* 49: 363–371.

———. 1909. "Über Raum, Zeit und Kraft als apriorische Formen der Mechanik." *Jahresbericht der Deutschen Mathematiker-Vereiningung* 18: 357–385.

———. 1927. "Die Axiome der Mechanik." In H. Geiger and K. Scheel (eds.), *Handbuch der Physik*, vol. 5, (*Grundlagen der Mechanik, Mechanik der Punkte und Starren Körper*). Berlin: Springer, 1–130.

Herglotz, Gustav. 1903. "Zur Elektronentheorie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 357–382.

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik in neuem Zusammenhänge dargestellt*. Leipzig: Barth.

———. 1956. *The Principles of Mechanics Presented in a New Form*. New York: Dover. English translation of (Hertz 1894).

Hertz, Paul. 1904. *Untersuchungen über unstetige Bewegungen eines Elektrons*. PhD Dissertation, Universität Göttingen.

Hessenberg, Gerhard. 1905. "Beweis des Desarguesschen Satzes aus dem Pascalschen." *Mathematische Annalen* 61, 161–172.
Hilbert, David. 1891. *Projective Geometry*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 535).
———. 1893–1894. *Die Grundlagen der Geometrie*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 541).
———. 1897. "Die Theorie der algebraischen Zahlkörper (Zahlbericht)." *Jahresbericht der Deutschen Mathematiker-Vereiningung* 4: 175–546. (Hilbert 1932–1935, vol. 1: 63–363.)
———. 1898–1899. *Mechanik*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 553).
———. 1899. *Grundlagen der Geometrie*. (*Festschrift zur Feier der Enthüllung des Gauss-Weber- Denkmals in Göttingen*.) Leipzig: Teubner.
———. 1901. "Mathematische Probleme." *Archiv für Mathematik und Physik* 1: 213–237, (Hilbert 1932–1935, vol. 3, 290–329).
———. 1902a. "Mathematical Problems." *Bulletin of the American Mathematical Society* 8: 437–479. English translation by M. Newson of (Hilbert 1901).
———. 1902b "Über die Grundlagen der Geometrie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 233–241. Reprinted in *Mathematische Annalen* 56, added as *Supplement IV* to (Hilbert 1903a).
———. 1902–1903. *Mechanik der Continua* I. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Instituts, Universität Göttingen, winter semester, 1902–1903, annotated by Berkovski).
———. 1903a. *Grundlagen der Geometrie,* (2nd revised edition – with five supplements). Leipzig: Teubner.
———. 1903b. *Mechanik der Continua II*. Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematisches Institut, Universität Göttingen. SS 1903, annotated by Berkovski.
———. 1904. "Über das Dirichletsche Prinzip." *Mathematische Annalen* 59: 161–186. (Hilbert 1932–1935, vol. 3: 15–37.) Reprinted from *Festschrift zur Feier des 150jährigen Bestehens der Königl. Gesellschaft der Wissenschaften zu Göttingen*, 1901.
———. 1905a. *Logische Principien des mathematischen Denkens*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Instituts, Universität Göttingen, summer semester, 1905, annotated by E. Hellinger.)
———. 1905b. *Logische Principien des mathematischen Denkens*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 558a) annotated by Max Born.
———. 1905c "Über die Grundlagen der Logik und der Arithmetik." In A. Kneser (ed.), *Verhandlungen aus der Dritten Internationalen Mathematiker-Kongresses in Heidelberg*, 1904, Teubner: Leipzig, 174-185. (English translation by G.B. Halsted: "On the Foundations of Logic and Arithmetic." *The Monist* 15: 338–352. Reprinted in van Heijenoort (ed.) 1967, 129–138.
———. 1905d. "Über das Dirichletsche Prinzip." *Journal für die reine und angewandte Mathematik* 129: 63–67. (Hilbert 1932–1935, vol. 3: 10–14. Reproduced from *Jahresbericht der Deutschen Mathematiker-Vereiningung* 8 (1900): 184–188.
———. 1906. *Mechanik der Kontinua*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, summer semester, 1906.)
———. 1910–1911. *Mechanik*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen winter semester, 1910–1911, annotated by F. Frankfurther.)
———. 1911–1912. *Kinetische Gastheorie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, winter semester, 1911–1912, annotated by E. Hecke.)
———. 1912a. *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Leipzig: Teubner.
———. 1912b. "Begründung der elementaren Strahlungstheorie." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 773–789; *Physikalische Zeitschrift* 13, 1056–1064, (Hilbert 1932–1935, vol. 3: 217–230).
———. 1912c. *Strahlungstheorie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen summer semester, 1912, annotated by E. Hecke.)
———. 1912–1913. *Molekulartheorie der Materie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, winter semester, 1912–1913.)
———. 1913a. "Begründung der elementaren Strahlungstheorie." *Jahresbericht der Deutschen Mathematiker-Vereiningung* 22: 1–20.
———. 1913b. *Elektronentheorie*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, summer semester, 1913.)
———. 1913c. *Elemente und Prinzipien der Mathematik*. (summer semester, 1913, Private Collection, Peter Damerow, Berlin.)
———. 1913–1914. *Elektromagnetische Schwingungen*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, winter semester, 1913–1914.)

———. 1916. *Die Grundlagen der Physik I*. (Manuscript/Typescript of Hilbert Lecture Notes. Bibliothek des Mathematischen Insitituts, Universität Göttingen, summer semester, 1913.)
———. 1918. "Axiomatisches Denken." *Mathematische Annalen* 78: 405–415. (Hilbert 1932–1935, vol. 3: 146–156. Reprinted in (Ewald 1999, Vol. 2, 1107–1115.)
———. 1930. "Naturerkennen und Logik." *Die Naturwissenschaften* 9: 59–63, (Hilbert 1932–1935, vol. 3, 378–387). English translation in (Ewald 1999, 1157–1165).
———. 1932–1935. *David Hilbert – Gesammelte Abhandlungen*, 3 vols. Berlin: Springer. (2nd ed. 1970).
———. 1992. *Natur und Mathematisches Erkennen: Vorlesungen, gehalten 1919–1920 in Göttingen. Nach der Ausarbeitung von Paul Bernays*. (Edited and with an English introduction by David E. Rowe.) Basel: Birkhäuser.
Hirosige, Tetu. 1976. "The Ether Problem, the Mechanistic Worldview, and the Origins of the Theory of Relativity." *Studies in History and Philosophy of Science* 7, 3–82.
Hochkirchen, Thomas H. 1999. *Die Axiomatisierung der Wahrscheinlichkeitsrechnung und ihre Kontexte. Von Hilberts sechstem Problem zu Kolmogoroffs Grundbegriffen*. Göttingen: Vandenhoeck & Ruprecht.
Hon, Giora. 1995. "Is the Identification of an Experimental Error Contextually Dependent? The Case of Kaufmann's Experiment and its Varied Reception." In J. Buchwald (ed.), *Scientific Practice: Theories and Stories of Doing Physics*. Chicago: Chicago University Press, 170–223.
Huntington, Edward V. 1904. "Set of Independent Postulates for the Algebra of Logic." *Trans. AMS* 5: 288–390.
Janssen, Michel. 2002. "Reconsidering a Scientific Revolution: The Case of Einstein versus Lorentz." *Physics in Perspective* 4: 421–446.
Jungnickel, Christa and Russel McCormmach. 1986. *Intellectual Mastery of Nature – Theoretical Physics form Ohm to Einstein*, 2 vols. Chicago: Chicago University Press.
Kirchhoff, Gustav. 1860. "Ueber das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermögen der Körper für Wärme und Licht. *Annalen der Physik* 109: 275–301.
Klein, Martin J. 1970. *Paul Ehrenfest: The Making of a Theoretical Physicist*. Amsterdam: North Holland.
Kragh, Helge. 1999. *Quantum Generations. A History of Physics in the Twentieth Century*. Princeton: Princeton University Press.
Kuhn, Thomas S. 1978. *Black-Body Theory and the Quantum Discontinuity, 1994–1912*. New York. Oxford University Press.
Lamb, Horace. 1895. *Hydrodynamics* (2nd ed.). Cambridge: Cambridge University Press.
Lanczos, Cornelius. 1962. *The Variational Principles of Mechanics*, (2nd ed.). Toronto: University of Toronto Press.
Larmor, Joseph. 1900. *Aether and Matter*. Cambridge: Cambridge University Press.
Lewis, Gilbert N. and Richard C. Tolman. 1909. "The Principle of Relativity, and Non-Newtonian Mechanics." *Philosophical Magazine* 18: 510–523.
Lorentz, Hendrik Antoon. 1895. *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern*. Leiden. In (Lorentz 1934–1939, vol. 5, 1–137).
———. 1898. "Die Fragen, welche die translatorische Bewegung des Lichtäthers betreffen." Verhandlungen GDNA 70 (2. Teil, 1. Hälfte), 56–65. In (Lorentz 1934–1939, vol. 7, 101–115).
———. 1900. "Considérations sur la Pesanteur." *Archives néerlandaises* 7 (1902), 325–338. Translated from *Versl. Kon. Akad. Wet. Amsterdam* 8: 325. In (Lorentz 1934–1939, vol. 5, 198–215).
———. 1904a. "Weiterbildung der Maxwellschen Theorie. Elektronentheorie." *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* 5, 2–14, 145–280.
———. 1904b. "Electromagnetic Phenomena in a System Moving with Velocity Smaller than that of Light." *Versl. Kon. Akad. Wet. Amsterdam* 6, 809–831. (Reprinted in A. Einstein et al. (1952) *The Principle of Relativity*. New York: Dover, 11–34.)
———. 1909. "Le partage de l'energie entre la matière pondérable et l'éther." In G. Castelnuovo (ed.) *Atti del IV congresso internazionale dei matematici* (Rome, 6–11 April 1909). Rome: Tipografia della R. Accademia dei Lincei, Vol. 1, 145–165. (Reprinted with revisions in *Nuovo Cimento* 16 (1908), 5–34.)
———. 1910. "Alte und neue Fragen der Physik." *Physikalische Zeitschrift* 11: 1234–1257. (*Collected Papers of Hendrik Anton Lorentz* 7, 205–207.)
———. 1934–1939. *Collected Papers of Hendrik Anton Lorentz* (9 vols.). The Hague: M. Nijhoff.
Lorey, Wilhelm. 1916. *Das Studium der Mathematik an den deutschen Universitäten seit Anfang des 19. Jahrhunderts*. Leipzig and Berlin: Teubner.
Love, Augustus E. H. 1901. "Hydrodynamik." In *Encyclopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* 4–3, 48–149.
Madelung, Erwin. 1912. "Die ponderomotorischen Kräfte zwischen Punktladungen in einem mit diffuser electromagnetischer Strahlung erfüllten Raume und die molekularen Kräfte." *Physikalische Zeitschrift* 13: 489–495.
McCormmach, Russel. 1970. "H. A. Lorentz and the Electromagnetic View of Nature." *Isis* 61: 457–497.

Mehrtens, Herbert. 1990. *Moderne - Sprache - Mathematik*. Frankfurt: Suhrkamp.
Miller, Arthur I. 1972. "On the Myth of Gauss's Experiment on the Physical Nature of Space." *Isis* 63: 345–348.
———. 1997. *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation, (1905–1911)*. New York: Springer.
Minkowski, Hermann. 1906. "Kapillarität." In *Encyclopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* V, 558–613.
———. 1907. *Wärmestrahlung*. Nachlass David Hilbert, (Cod. Ms. D. Hilbert, 707).
Montgomery, Deane and Leo Zippin. 1952. "Small groups of finite-dimensional groups." *Annals of Mathematics* 56: 213–241.
Moore, Eliahim H. 1902. "Projective Axioms of Geometry." *Trans. American Mathematical Society* 3: 142–158.
Moore, Gregory H. 1987. "A House Divided Against Itself: the Emergence of First-Order Logic as the Basis for Mathematics." In E. R. Phillips (ed.), *Studies in the History of Mathematics*, MAA Studies in Mathematics, 98–136.
———. 1995. "The Axiomatization of Linear Algebra: 1875–1940." *Historia Mathematica* 22: 262–303.
Neumann, Carl Gottfried. 1870. *Ueber die Principien der Galilei-Newton'schen Theorie*. Leipzig: Teubner.
Nernst, Walter. 1906. "Ueber die Berechnung chemischer Gleichgewichte aus thermischen Messungen." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen* 1: 1–39.
Noll, Walter. 1959. "The Foundations of Classical Mechanics in the Light of Recent Advances in Continuum Mechanics." In *The Axiomatic Method with Special Reference to Geometry and Physics*. Amsterdam: North Holland, 266–281. (Reprinted in W. Noll, *The Foundations of Mechanics and Thermodynamics*, New York/ Heidelberg/ Berlin: Springer (1974), 32–47.)
North, John. 1965. *The Measure of the Universe*. Oxford: Clarendon Press.
Peckhaus, Volker. 1990. *Hilbertprogramm und Kritische Philosophie. Der Göttinger Modell interdisziplinärer Zusammenarbeit zwischen Mathematik und Philosophie*. Göttingen: Vandenhoeck & Ruprecht.
Peckhaus, Volker and Reinhard Kahle. 2002. "Hilbert's Paradox." *Historia Mathematica* 29: 157–175.
Planck, Max. 1899. "Über irreversible Strahlungsvorgänge. Dritte Mitteilung (Schluss)." *Königlich Preussische Akademie der Wissenschaften (Berlin) Sitzungsberichte*, 440–480.
———. 1900. "Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum." *Verhandlungen der Deutsche Physikalische Gesellschaft* 2: 237–245.
———. 1906. *Vorlesungen über die Theoire der Wärmestrahlung*, Leipzig.
Planck, Max et al. 1914. *Vorträge über die kinetische Theorie der Materie und der Elektrizität. Gehalten in Göttingen auf Einladung der Kommission der Wolfskehlstiftung*. Leipzig and Berlin: Teubner.
Poincaré, Henri. 1901. *Electricité et optique: La lumière et les théories électrodynamiques. Leçons professées à la Sorbonne en 1999, 1990, et 1990,* (eds. J. Blondin and E. Néculcéa). Paris: **VERLAG**?.
———. 1912. "Sur la théorie des quanta." *Journal Phys. Théor. et Appl*. 2: 5–34.
Prandtl, Ludwig. 1904. "Über Flüssigkeitbewegung bei sehr kleiner Reibung." In A. Kneser (ed.), *Verhandlungen aus der Dritten Internationalen Mathematiker-Kongresses in Heidelberg, 1904*. Teubner: Leipzig, 484–491.
Pyenson, Lewis R. 1979. "Physics in the Shadows of Mathematics: The Göttingen Electron-Theory Seminar of 1905." *Archive for History of Exact Sciences* 21: 55–89. (Reprinted in Pyenson 1985, 101–136.)
———. 1982. "Relativity in Late Wilhelmian Germany: The Appeal to a Pre-established Harmony Between Mathematics and Physics." *Archive for History of Exact Sciences* 138–155. (Reprinted in Pyenson 1985, 137–157.)
———. 1985. *The Young Einstein - The Advent of Relativity.* Bristol and Boston: Adam Hilger Ltd.
Rausenberg, O. 1988. *Lehrbuch der analytischen Mechanik*. Erster Band*: Mechanik der materiellen Punkte*. Zweiter Band: *Mechanik der zusammenhängenden Körper*. Leipzig, Teubner.
Reid, Constance. 1970. *Hilbert*. Berlin/New York: Springer.
Reiff, R. 1900. "Die Druckkräfte in der Hydrodynamik und die Hertzsche Mechanik." *Annalen der Physik* 1: 225–231.
Rowe, David E. 1996. "I 23 problemi de Hilbert: la matematica agli albori di un nuovo secolo." *Storia del XX Secolo: Matematica-Logica-Informatica*. Rome: Enciclopedia Italiana.
———. 1999. "Perspective on Hilbert" (Review of Mehrtens 1990, Peckhaus 1990, and Toepell 1986), *Perspectives on Science*, 5 (4): 533–570.
Rüdenberg, Lily and Hans Zassenhaus (eds.). 1973. *Hermann Minkowski - Briefe an David Hilbert*. Berlin/New York: Springer.
Sackur, Otto. 1912. *Lehrbuch der Thermochemie und Thermodynamik*. Berlin: Springer.
Scanlan, Michael. 1991. "Who were the American Postulate Theorists?" *Journal of Symbolic Logic* 56: 981–1002.

Schirrmacher, Arne. 2003. "Experimenting Theory: The Proofs of Kirchhoff's Radiation Law before and after Planck." *Historical Studies in the Physical Sciences* 33 (2): 299–335.

Schmidt, Arnold. 1933. "Zu Hilberts Grundlegung der Geometrie," (Hilbert 1932–1935, Vol. 2: 404–414).

Scholz, Erhard. 1992. "Gauss und die Begründung der 'höhere' Geodäsie." In M. Folkerts et al. (eds.), *Amphora Festschrift für Hans Wussing zu seinem 65. Geburtstag*. Berlin: Birkhäuser, 631–648.

———. 2004. "C. F. Gauß' Präzisionsmessungen terrestrischer Dreicke und seine Überlegungen zur empirischen Fundierung der Geometrie in den 1820er Jahren." In R. Seising, M. Folkerts, and U. Hashagen (eds.), *Form, Zahl, Ordnung: Studien zur Wissenschafts- und Technikgeschichte: Ivo Schneider zum 65. Geburtstag*. Wiesbaden: Franz Steiner Verlag.

Schur, Friedrich. 1898. "Über den Fundamentalsatz der projektiven Geometrie." *Mathematische Annalen* 51: 401–409.

———. 1901. "Über die Grundlagen der Geometrie." *Mathematische Annalen* 55: 265–292.

———. 1903. "Über die Zusammensetzung von Vektoren." *Zeitschrift für Mathematik und Physik* 49: 352–361.

Schwarzschild, Karl. 1903. "Zur Elektrodynamik: III. Ueber die Bewegung des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 245–278.

Schimmack, R. 1903. "Ueber die axiomatische Begründung der Vektoraddition." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse,* 317–325.

Sinaceur, Hourya. 1984. "De D. Hilbert á E. Artin: les différents aspects du dix-septième problème et les filiations conceptuelles de la théorie des corps reels clos." *Archive for History of Exact Sciences* 29: 267–287.

———. 1991. *Corps et Modèles*. Paris: Vrin.

Sommerfeld, Arnold. 1904a. "Zur Elektronentheorie: I. Allgemeine Untersuchung des Feldes eines beliebig bewegten Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 99–130.

———1904b. "Zur Elektronentheorie: II. Grundlagen für eine allgemeine Dynamik des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 363–439.

———. 1905. "Zur Elektronentheorie: III. Ueber Lichtgeschwindichkeits- und Ueberlichtgeschwindichkeits-Elektronen." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 99–130; "II. Grundlagen für eine allgemeine Dynamik des Elektrons." *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematische-Physikalische Klasse*, 201–235.

Thiele, Rüdiger. 2003. "Hilbert's Twenty-Fourth Problem." *American Mathematical Monthly* 110 (1): 1–24.

Toepell, Michael M. 1986. *Über die Entstehung von David Hilberts 'Grundlagen der Geometrie'*. Göttingen: Vandenhoeck & Ruprecht.

Torretti, Roberto. 1978. *Philosophy of Geometry from Riemann to Poincaré*. Dordrecht: Reidel.

Truesdell, Clifford. 1968. *Essays in the History of Mechanics*. New York: Springer.

Veblen, Oswald. 1904. "A System of Axioms for Geometry." *Trans. American Mathematical Society* 5: 343–384.

Voss, Aurel. 1901. "Die Principien der rationellen Mechanik." *Encyclopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* IV–1, 3–121.

Warwick, Andrew C. 1991. "On the Role of the FizGerald-Lorentz Contraction Hypothesis in the Development of Joseph Larmo's Theory of Matter." *Archive for History of Exact Sciences* 43: 29–91.

———. 2003. *Masters of Theory. Cambridge and the Rise of Mathematical Physics*. Chicago: The University of Chicago Press.

Weyl, Hermann. 1944. "David Hilbert and his Mathematical Work." *Bull. American Mathematical Society* 50: 612–654.

Wiechert, Emil. 1899. *Grundlagen der Elektrodynamik. Festschrift zur Feier der Enthüllung des Gauß-Weber-Denkmals in Göttingen*. Leipzig: Teubner.

———. 1901. "Elektrodymanische Elementargesetze. "*Annalen der Physik* 4: 667–689.

Wien, Wilhelm. 1900. "Ueber die Möglichkeit einer elektromagnetischen Begründung der Mechanik." Archives néerlandaises 5: 96–104. (Reprinted in *Phys. Chem. Ann.* 5 (1901), 501–513.)

Yavetz, Ido. 1995. *From Obscurity to Enigma. The Work of Oliver Heaviside, 1872–1889*. Boston: Birkhäuser. (*Science Networks*, Vol. 16.)

JÜRGEN RENN AND JOHN STACHEL

HILBERT'S FOUNDATION OF PHYSICS: FROM A THEORY OF EVERYTHING TO A CONSTITUENT OF GENERAL RELATIVITY

1. ON THE COMING INTO BEING AND FADING AWAY OF AN ALTERNATIVE POINT OF VIEW

1.1 The Legend of a Royal Road to General Relativity

Hilbert is commonly seen as having publicly presented the derivation of the field equations of general relativity on 20 November 1915, five days before Einstein and after only half a year's work on the subject in contrast to Einstein's eight years of hardship from 1907 to 1915.[1] We thus read in Kip Thorne's fascinating account of recent developments in general relativity (Thorne 1994, 117):

> Remarkably, Einstein was not the first to discover the correct form of the law of warpage [of space-time, i.e. the gravitational field equations], the form that obeys his relativity principle. Recognition for the first discovery must go to Hilbert. In autumn 1915, even as Einstein was struggling toward the right law, making mathematical mistake after mistake, Hilbert was mulling over the things he had learned from Einstein's summer visit to Göttingen. While he was on an autumn vacation on the island of Rugen in the Baltic the key idea came to him, and within a few weeks he had the right law–derived not by the arduous trial-and-error path of Einstein, but by an elegant, succinct mathematical route. Hilbert presented his derivation and the resulting law at a meeting of the Royal Academy of Sciences in Göttingen on 20 November 1915, just five days before Einstein's presentation of the same law at the Prussian Academy meeting in Berlin.[2]

Hilbert himself emphasized that he had two separate starting points for his approach: Mie's electromagnetic theory of matter as well as Einstein's attempt to base a theory of gravitation on the metric tensor. Hilbert's superior mastery of mathematics apparently allowed him to arrive quickly and independently at combined field equa-

1 For discussions of Einstein's path to general relativity see (Norton 1984; Renn and Sauer 1999; Stachel 2002), "The First Two Acts", "Pathways out of Classical Physics …", and "Untying the Knot …", (in vols. 1 and 2 of this series). For historical reviews of Hilbert's contribution, see (Guth 1970; Mehra 1974; Earman and Glymour 1978; Pais 1982, 257–261; Corry 1997; 1999a; 1999b; 1999c; Corry, Renn, and Stachel 1997; Stachel 1989; 2002; Sauer 1999; 2002), "The Origin of Hilbert's Axiomatic Method …" and "Einstein Equations and Hilbert Action" (both in this volume).

2 For a similar account see (Fölsing 1997, 375–376).

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

tions for the electromagnetic and gravitational fields. Although his use of Mie's ideas initially led Hilbert to a theory that was, from the point of view of the subsequent general theory of relativity, restricted to a particular source for the gravitational field—the electromagnetic field—he is nevertheless regarded by many historians of science and physicists as the first to have established a mathematical framework for general relativity that provides both essential results of the theory, such as the field equations, and a clarification of previously obscure conceptual issues, such as the nature of causality in generally-covariant field theories.[3] His contributions to general relativity, although initially inspired by Mie and Einstein, hence appear as a unique and independent achievement. In addition, Hilbert is seen by some historians of science as initiating the subsequent search for unified field theories of gravitation and electromagnetism.[4] In view of all these results, established within a very short time, it appears that Hilbert indeed had found an independent "royal road" to general relativity and beyond.

In a recent paper with Leo Corry, we have shown that Hilbert actually did not anticipate Einstein in presenting the field equations (Corry, Renn, and Stachel 1997).[5] Our argument is based on the analysis of a set of proofs of Hilbert's first paper,[6] hereafter referred to as the "Proofs". These Proofs not only do not include the explicit form of the field equations of general relativity, but they also show the original version of Hilbert's theory to be in many ways closer to the earlier, non-covariant versions of Einstein's theory of gravitation than to general relativity. It was only *after* the publication on 2 December 1915 of Einstein's definitive paper that Hilbert modified his theory in such a way that his results were in accord with those of Einstein.[7] The final version of his first paper, which was not published until March 1916, now includes the explicit field equations and has no restriction on general covariance (Hilbert 1916).[8] Hilbert's second paper, a sequel to his first communication, in which he first discussed causality, apparently also underwent a major revision before eventually being published in 1917 (Hilbert 1917).[9]

---

3 See (Howard and Norton 1993).

4 See, for example, (Vizgin 1989), who refers to "Hilbert's 1915 unified field theory, in which the attempt was first made to unite gravitation and electromagnetism on the basis of the general theory of relativity" (see p. 301).

5 See also (Stachel 1999), reprinted in (Stachel 2002).

6 A copy of the proofs of Hilbert's first paper is preserved at Göttingen, in SUB Cod. Ms. 634. They comprise 13 pages and are virtually complete, apart from the fact that roughly the upper quarter of two pages (7 and 8) is cut off. The Proofs are dated "submitted on 20 November 1915." The Göttingen copy bears a printer's stamp dated 6 December 1915 and is marked in Hilbert's own hand "First proofs of my first note." In addition, they carry several marginal notes in Hilbert's hand, which are discussed below. A complete translation of the Proofs is given in this volume.

7 The conclusive paper is (Einstein 1915e), which Hilbert lists in the references in (Hilbert 1916).

8 In the following referred to as Paper 1.

9 In the following referred to as Paper 2.

1.2 The Transformation of the Meaning of Hilbert's Work

Hilbert presented his contribution as emerging from a research program that was entirely his own—the search for an axiomatization of physics as a whole—creating a synthesis of electromagnetism and gravitation. This view of his achievement was shared by Felix Klein, who took the distinctiveness of Hilbert's approach as an argument against seeing it from the perspective of a priority competition with Einstein:

> There can be no talk of a priority question in this connection, since both authors are pursuing quite different trains of thought (and indeed, so that initially the compatibility of their results did not even seem certain). Einstein proceeds *inductively* and immediately considers arbitrary material systems. Hilbert *deduces* from previously postulated basic variational principles, while he additionally allows the restriction to electrodynamics. In this connection, Hilbert was particularly close to Mie.[10]

It is clear that, even if one disregards the non-covariant version of his theory as presented in the proofs version of his first paper, both Hilbert's original programmatic aims as well as the interpretation he gave of his own results do not fit into the framework of general relativity as we understand it today. To give one example, which we shall discuss in detail below: In the context of Hilbert's attempt at a synthesis of electromagnetism and gravitation theory, he interpreted the contracted Bianchi identities as a substitute for the fundamental equations of electromagnetism, an interpretation that was soon recognized to be problematic by Hilbert himself.

With hindsight, however, there can be little doubt that a number of important contributions to the development of general relativity do have roots in Hilbert's work: For instance, not so much the variational formulation of the gravitational field equations, an idea which had already been introduced by Einstein[11]; but the choice of the Ricci scalar as the gravitational term in this Lagrangian; and the first hints of Noether's theorem.

The intrinsic plausibility of each of these two perspectives: viewing Hilbert's work as either aiming at a theory differing from general relativity, or as a contribution to general relativity, represents a puzzle. How can Hilbert's contributions be interpreted as making sense only within an independent research program, different in essence from that of Einstein, if ultimately they came to be seen, at least by most physicists, as constituents of general relativity? This puzzle raises a profound historical question concerning the nature of scientific development: how were Hilbert's results, produced within a research program originally aiming at an electrodynamic

---

10 "Von einer Prioritätsfrage kann dabei keine Rede sein, weil beide Autoren ganz verschiedene Gedankengänge verfolgen (und zwar so, daß die Verträglichkeit der Resultate zunächst nicht einmal sicher schien). Einstein geht *induktiv* vor und denkt gleich an beliebige materielle Systeme. Hilbert *deduziert*, indem er übrigens die [...] Beschränkung auf Elektrodynamik eintreten läßt, aus voraufgestellten obersten Variationsprinzipien. Hilbert hat dabei insbesondere auch an Mie angeknüpft." (Klein 1921, 566). The text was originally published in 1917; see (Klein 1917). The quote is from a footnote to remarks added to the 1921 republication. For a recent reconstruction of Hilbert's perspective, see (Sauer 1999).

11 See "Untying the Knot …" (in vol. 2 of this series).

foundation for *all* of physics, eventually transformed into constituents of general relativity, a theory of gravitation? The pursuit of this question promises insights into the processes by which scientific results acquire and change their meaning and, in particular, into the process by which a viewpoint that is different from the one eventually accepted as mainstream emerges and eventually fades away.[12]

Hilbert's work on the foundations of physics turns out to be especially suited for such an analysis, not only because the proofs version of his first paper provides us with a previously unknown point of departure for following his development, but also because he came back time and again to these papers, rewriting them in terms of the insights he had meanwhile acquired and in the light of the developments of Einstein's "mainstream" program. In this paper we shall interpret Hilbert's revisions as indications of the conceptual transformation that his original approach underwent as a consequence of the establishment and further development of general relativity by Einstein, Schwarzschild, Klein, Weyl, and others, including Hilbert himself. We will also show that Hilbert's own understanding of scientific progress induced him to perceive this transformation as merely an elimination of errors and the introduction of improvements and elaborations of a program he had been following from the beginning.

1.3 Structure of the Paper

In the *second section* of this paper ("The origins of Hilbert's program in the 'nostrification' of two speculative physical theories"), we shall analyze the emergence of Hilbert's program for the foundations of physics from his attempt to synthesize, in the form of an axiomatic system, techniques and results of Einstein's 1913/14 non-covariant theory of gravitation and Mie's electromagnetic theory of matter. It will become clear that Hilbert's research agenda was shaped in large part by his understanding of the axiomatic formulation of physical theories, by the technical problems of achieving the synthesis of these two theories, and by open problems in Einstein's theory.

In the *third section* ("Hilbert's attempt at a theory of everything: the proofs of his first paper"), we shall interpret the proofs version of Hilbert's first paper as an attempt to realize the research program reconstructed in the second section. In particular, we shall show that, in the course of pursuing this program, he abandoned his original goal of founding all of physics on electrodynamics, now treating the gravitational field as more fundamental. We shall argue that this reversal was induced by mathematical results, to which Hilbert gave a problematic physical interpretation suggested by his research program; and that the mathematical result at the core of Hilbert's attempt to establish a connection between gravitation and electromagnetism originated in Einstein's claim of 1913/14 that generally-covariant field equations are not compatible with physical causality, a claim supported by Einstein's well-known "hole-argument." Hilbert thus turned Einstein's argument against general covariance into support for Hilbert's own attempt at a unified theory of gravitation and electro-

12 Cf. (Stachel 1994).

magnetism. Hilbert also followed Einstein's 1913/14 attempt to relate the existence of a preferred class of coordinate systems to the requirement of energy conservation. Hilbert's definition of energy, however, was not guided by Einstein's but rather by the goal of establishing a link with Mie's theory. Hilbert's unified theory thus emerges as an extension of Einstein's non-covariant theory of gravitation, in which Mie's speculative theory of matter plays the role of a touchstone, a role played for Einstein by the principle of energy-momentum conservation in classical and special relativistic physics and in Newton's theory of gravitation.

In the *fourth section* ("Hilbert's physics and Einstein's mathematics: the exchange of late 1915") we shall examine Hilbert's and Einstein's exchange of letters at the end of 1915, focussing on the ways in which they mutually influenced each other. We show that Hilbert's attempt at combining a theory of gravitation with a theory of matter had an important impact on the final phase of Einstein's work. Hilbert's vision, which Einstein temporarily adopted, provided the latter with a rather exotic perspective but allowed him to obtain a crucial result, the calculation of Mercury's perihelion precession. This, in turn, guided his completion of the general theory of relativity, but at the same time rendered obsolete its grounding in a specific theory of matter. For Hilbert's theory, on the other hand, Einstein's conclusive paper on general relativity represented a major challenge. It undermined the entire architecture; in particular, the connections Hilbert saw between energy conservation, causality, and the need for a restriction of general covariance.

In the *fifth section* ("Hilbert's adaptation of his theory to Einstein's results: the published versions of his first paper") we shall first discuss how, under the impact of Einstein's results in November 1915, Hilbert modified essential elements of his theory before its publication in March 1916. He abandoned the attempt to develop a non-covariant theory, without as yet having found a satisfactory solution to the causality problem that Einstein had previously raised for generally-covariant theories. He replaced his original, non-covariant notion of energy by a new formulation, still differing from that of Einstein and mainly intended to strengthen the link between his own theory and Mie's electrodynamics. In fact, Hilbert did not abandon his aim of providing a foundation for all of physics. He still hoped to construct a field-theoretical model of the electron and derive its laws of motion in the atom, without, however, getting far enough to include any results in his paper. His first paper was republished twice, in 1924 and 1933, each time with significant revisions. We shall show that Hilbert eventually adopted the understanding of energy-momentum conservation developed in general relativity, thus transforming his ambitious program into an application of general relativity to a special kind of source, matter as described by Mie's theory.

In the *sixth section* ("Hilbert's adoption of Einstein's program: the second paper and its revisions") we shall show that Hilbert's second paper, published in 1917, is the outcome of his attempt to tackle the unsolved problems of his theory in the light of Einstein's results, in particular the causality problem; and at the same time to keep up with the rapid progress of general relativity. In fact, instead of pursuing the conse-

quences of his approach for microphysics, as he originally intended, he now turned to solutions of the gravitational field equations, relating them to the mathematical tradition inaugurated by Gauss and Riemann of exploring the applicability of Euclidean geometry to the physical world. In this way, he effectively worked within the program of general relativity and contributed to solving such problems as the uniqueness of the Minkowski solution and the derivation of the Schwarzschild solution; but he was less successful in dealing with the problem of causality in a generally-covariant theory. Although he followed Einstein in focussing on the invariant features of such a theory, he attempted to develop his own solution to the causality problem, different from that of Einstein. Whereas Einstein resolved the ambiguities he had earlier encountered in the hole argument by the insight that in general relativity coordinate systems have no physical significance apart from the metric, Hilbert attempted to find a purely "mathematical response" to this problem, formulating the causality condition in terms of the Cauchy or initial-value problem for the generally-covariant field equations. While it initiated an important line of research in general relativity, this first attempt not only failed to incorporate Einstein's insights into the physical interpretation of general relativity but also suffered from Hilbert's inadequate treatment of the Cauchy problem for such a theory, a treatment that was finally corrected by the editors of the revised version published in 1933.

In the *seventh section* ("The fading away of Hilbert's point of view in the physics and mathematics communities") we shall analyze the reception of Hilbert's work in contemporary literature on general relativity and unified field theories, as well as its later fate in the textbook tradition. We show that, in spite of Hilbert's emphasis on the distinctiveness of his approach, his work was perceived almost exclusively as a contribution to general relativity. It will become clear that this reception was shaped largely by the treatment of Hilbert's work in the publications of Einstein and Weyl, although, by revising his own contributions in the light of the progress of general relativity, Hilbert was not far behind in contributing to the complete disappearance of his original, distinctive point of view. This disappearance had two remarkable consequences: First, deviations of Hilbert's theory from general relativity, such as his interpretation of the contracted Bianchi identities as the coupling between gravitation and electromagnetism, went practically unremarked. Second, in spite of his attempt to depict himself as the founding father of unified field theories, the early workers in this field tended to ignore his contribution, denying him a prominent place in their intellectual ancestry. Instead, Hilbert was assigned a prominent place in the history of general relativity, even ascribing to him achievements that were not his, such as the first formulation of the field equations or the complete clarification of the question of causality. The ease with which his work could be assimilated to general relativity provides further evidence of a different kind for the tenuous and unstable character of his own framework.

In the *eighth and final section* ("At the end of a royal road") we shall compare Hilbert's and Einstein's approaches in an effort to understand Hilbert's gradual rapprochement with general relativity. Einstein had followed a double strategy in creat-

ing general relativity: trying to explore the mathematical consequences of physical principles on the one hand; and systematically checking the physical interpretation of mathematical results, on the other. Hilbert's initial approach encompassed a much narrower physical basis. Starting from a few problematic physical assumptions, Hilbert elaborated a mathematically complex framework, but never succeeded in finding any concrete physical consequences of this framework other than those that had been or could be found within Einstein's theory of general relativity. Nevertheless, Hilbert's assimilation of specific results from the mainstream tradition of general relativity into his framework eventually changed the character of this framework, transforming his results into contributions to general relativity. Thus, in a sense, Hilbert's assimilation of insights from general relativity served as a substitute for the physical component of Einstein's double strategy that was originally lacking in Hilbert's own approach. So this double strategy emerges not only as a successful heuristic characterizing Einstein's individual pathway, but as a particular aspect of the more general process by which additional knowledge was integrated into the further development of general relativity.

2. THE ORIGINS OF HILBERT'S PROGRAM IN THE "NOSTRIFICATION" OF TWO SPECULATIVE PHYSICAL THEORIES

Leo Corry has explored in depth the roots and the history of Hilbert's program of axiomatization of physics and, in particular, its impact on his 1916 paper *Foundations of Physics*.[13] We can therefore limit ourselves to recapitulating briefly some essential elements of this program. Hilbert conceived of the axiomation of physics not as a definite foundation that has to *precede* empirical research and theory formation, but as a *post-hoc* reflection on the results of such investigations with the aim of clarifying the logical and epistemological structure of the assumptions, definitions, etc., on which they are built.[14] Nevertheless, Hilbert expected that a proper axiomatic foundation of physics would not be shaken every time a new empirical fact is discovered; but rather that new, significant facts could be incorporated into the existing body of knowledge without changing its logical structure. Furthermore, Hilbert expected that, rather than emerging from the reorganization of the existing body of knowledge, the concepts used in an axiomatic foundation of physics should be those already familiar from the history of physics. Finally, Hilbert was convinced that one can distinguish sharply between the particular, empirical and the universal ingredients of a physical theory.

Accordingly, the task that Hilbert set for himself was not to find new concepts serving to integrate the existing body of physical knowledge into a coherent conceptual whole, but rather to formulate appropriate axioms involving the already-existing

13 See (Corry 1997; 1999a; 1999b; 1999c; see also Sauer 1999, section 1) and "The Origin of Hilbert's Axiomatic Method …" (in this volume).

14 For evidence of the following claims, see, in particular, Hilbert's lecture notes (Hilbert 1905; 1913), extensively discussed in Corry's papers.

physical concepts; axioms which allow the reconstruction of available physical knowledge by deduction from these axioms. Consequently, his interest in the axiomatization of physics was oriented toward the reductionist attempts to found all of physics on the basis of either mechanics or electrodynamics (the mechanical or electromagnetic worldview). Indeed, in his discussions of the foundations of physics before 1905, the axiomatization of mechanics was central; while, at some point after the advent of the special theory of relativity, Hilbert now placed his hopes in an axiomatization of all physics based on electrodynamics.[15] In spite of the conceptual revolution brought about by special relativity, involving not only the revision of the concepts of space and time but also the autonomy of the field concept from that of the aether, Hilbert nevertheless continued to rely on traditional concepts such as force and rigidity as the building blocks for his axiomatization program.[16]

An axiomatic synthesis of existing knowledge such as that pursued by Hilbert in physics apparently also had a strategic significance for Göttingen mathematicians making it possible for them to leave their distinctive mark on a broad array of domains, which were thus "appropriated," not only intellectually but also in the sense of professional responsibility for them. Minkowski's attempt to present his work on special relativity as a decisive mathematical synthesis of the work of his predecessors may serve as an example.[17] Discussing an accusation that Emmy Noether had neglected to acknowledge her intellectual debt to British and American algebraists, Garrett Birkhoff wrote:

> This seems like an example of German 'nostrification:' reformulating other people's best ideas with increased sharpness and generality, and from then on citing the local reformulation.[18]

2.1 Mie's Theory of Matter

By 1913, Hilbert expected that the electron theory of matter would provide the foundation for all of physics. It is therefore not surprising to find him shortly afterwards attracted to Mie's theory of matter, a non-linear generalization of Maxwell's electrodynamics that aimed at the overcoming of the dualism between "aether" and "ponderable matter." Indeed, Mie had introduced a generalized Hamiltonian formalism for electrodynamics, allowing for non-linear couplings between the field variables, in the hope of deriving the electromagnetic properties of the "aether" as well as the particulate structure of matter from one and the same variational principle.[19] Mie's theory thus not only corresponded to Hilbert's hope to found all of physics on the concepts

15 For a discussion of Hilbert's turn from mechanical to electromagnetic monism, see (Corry 1999a, 511–517).

16 See (Hilbert 1913, 13).

17 This attempt is extensively discussed in (Walter 1999). See also (Rowe 1989).

18 Garrett Birkhoff to Bartel Leendert van der Waerden, 1 November 1973 (Eidgenössische Technische Hochschule Zürich, Handschriftenabteilung, Hs 652:1056); quoted from (Siegmund-Schultze 1998, 270). We thank Leo Corry for drawing our attention to this letter.

of electrodynamics; but it must also have been attractive to him because it was based upon the variational calculus, a tool, with the usefulness of which for the axiomatization of physical theories Hilbert was quite familiar.[20] However, Mie's theory was far from able to provide specific results concerning the electromagnetic properties of matter, results which could be confronted with empirical data. Rather, the theory provides only a framework; a suitable "world function" (Lagrangian) must still be found, from which such concrete predictions may then be derived. Mie gave examples of such world functions that, however, were meant to be no more than illustrations of certain features of his framework. In fact, Mie could not have considered these examples as the basis of a specific physical theory since they are not even compatible with basic features of physical reality such as the existence of an elementary quantum of electricity. Concerning his principal example, later taken up by Hilbert, Mie himself remarked:

> A world that is governed by the world function
>
> $$\phi = -\frac{1}{2}\eta^2 + \frac{1}{6}a \cdot \chi^6 \tag{1}$$
>
> must ultimately agglomerate into two large lumps of electric charges, one positive and one negative, and both these lumps must continually tend to separate further and further from each other.[21]

Mie drew the obvious conclusion that the unknown world function he eventually hoped to find must be more complicated than this and the other examples he had considered.[22]

Hilbert based his work on a formulation of Mie's framework actually due to Max Born.[23] In a paper of 1914, Born showed that Mie's variational principle can be considered as a special case of a four-dimensional variational principle for the deformation of a four-dimensional continuum involving the integral:[24]

$$\int \phi(a_{11}, a_{12}, a_{13}, a_{14}; a_{21} \ldots; u_1, \ldots u_4) dx_1 dx_2 dx_3 dx_4 \ . \tag{2}$$

---

19 Mie's theory was published in three installments: (Mie 1912a; 1912b; 1913). For a concise account of Mie's theory, see (Corry 1999b), see also the Editorial Note in this volume. In the recent literature on Mie's theory, the problematic physical content of this theory (and hence of its adaptation by Hilbert) plays only a minor role; see the discussion below.

20 See, in particular, (Hilbert 1905).

21 "Eine Welt, die durch die Weltfunktion (1) regiert würde, müßte sich also schließlich zu zwei großen Klumpen elektrischer Ladungen zusammenballen, einem positiven und einem negativen, und diese beiden Klumpen müßten immer weiter und weiter voneinander wegstreben." (Mie 1912b, 38) For the meaning of Mie's formula and its ingredients in Hilbert's version, see (33) below.

22 See (Mie 1912b, 40).

23 For a discussion of Born's role as Hilbert's informant about both Mie's and Einstein's theories, see (Sauer 1999, 538–539).

24 See (Born 1914).

Here ϕ is a Lorentz scalar, and:

$$u_\alpha = u_\alpha(x_1, x_2, x_3, x_4) \qquad \alpha = 1, \ldots 4 \tag{3}$$

are the projections onto four orthogonal axes of the displacements of the points of the four-dimensional continuum from their equilibrium positions regarded as functions of the quasi-Cartesian coordinates x_1, x_2, x_3, x_4 along these axes, and

$$a_{\alpha\beta} = \frac{\partial u_\alpha}{\partial x_\beta} \tag{4}$$

are their derivatives. Furthermore, Born showed that the characteristic feature of Mie's theory lies in the ansatz that the function ϕ depends only on the antisymmetric part of $a_{\alpha\beta}$:

$$a_{\alpha\beta} - a_{\beta\alpha} = \frac{\partial u_\alpha}{\partial x_\beta} - \frac{\partial u_\beta}{\partial x_\alpha}. \tag{5}$$

Mie's four-dimensional continuum could thus be regarded as a four-dimensional spacetime generalization of MacCullagh's three-dimensional aether. MacCullagh had derived equations corresponding to Maxwell's equations for stationary electrodynamic processes from the assumption that the vortices of the aether, rather than its deformations, store its energy.[25]

What role does gravitation play in Mie's theory? Mie opened the series of papers on his theory with a programmatic formulation of his goals, among them to establish a link between the existence of matter and gravitation:

> The immediate goals that I set myself are: to explain the existence of the indivisible electron and: to view the actuality of gravitation as in a necessary connection with the existence of matter. I believe one must start with this, for electric and gravitational effects are surely the most direct expression of those forces upon which rests the very existence of matter. It would be senseless to imagine matter whose smallest parts did not possess electric charges, equally senseless however matter without gravitation.[26]

Initially Mie hoped that he could explain gravitation on the basis of his non-linear electrodynamics alone, without introducing further variables. His search for a new theory of gravitation was guided by a simple model, according to which gravitation is a kind of "atmosphere," arising from the electromagnetic interactions inside the atom:

> An atom is an agglomeration of a larger number of electrons glued together by a relatively dilute charge of opposite sign. Atoms are probably surrounded by more substantial

25 See (Whittaker 1951, 142–145, Schaffner 1972, 59–68).

26 "Die nächsten Ziele, die ich mir gesteckt habe, sind: die Existenz des unteilbaren Elektrons zu erklären und: die Tatsache der Gravitation mit der Existenz der Materie in einem notwendigen Zusammenhang zu sehen. Ich glaube, daß man hiermit beginnen muß, denn die elektrischen und die Gravitationswirkungen sind sicher die unmittelbarsten Äußerungen der Kräfte, auf denen die Existenz der Materie überhaupt beruht. Es wäre sinnlos, Materie zu denken, deren kleinste Teilchen nicht elektrische Ladungen haben, ebenso sinnlos aber Materie ohne Gravitation." See (Mie 1912a, 511–512).

> atmospheres, which however are still so dilute that they do not cause noticeable electric fields, but which presumably are asserted in gravitational effects.[27]

In his third and conclusive paper, however, he explicitly withdrew this model and was forced to introduce the gravitational potential as an additional variable.[28] There is thus no intrinsic connection between gravitation and the other fields in Mie's theory. By representing gravitation as an additional term in his Lagrangian giving rise to a four-vector representation of the gravitational field, he effectively returned to Abraham's gravitation theory which he had earlier rejected.[29] As a consequence, his treatment of gravitation suffers from the same objections that were raised in contemporary discussions of Abraham's theory. In summary, Mie's theory of gravitation was far from reaching the goals he had earlier set for it.

2.2 Einstein's Non-Covariant "Entwurf" Theory of Gravitation

In 1915, Hilbert became interested in Einstein's theory of gravitation after a series of talks on this topic by Einstein between 28 June and 5 July of that year in Göttingen.[30] Hilbert's attraction to Einstein's approach may have stemmed from his dissatisfaction with the contrast between Mie's programmatic statements about the need for a unification of gravitation and electromagnetism and the unsatisfactory treatment of gravitation in Mie's actual theory. This may well have motivated Hilbert to look at other theories of gravitation and perhaps even to invite Einstein. But apart from the shortcomings of Mie's theory, Hilbert's fascination with Einstein's approach to gravitation probably is rooted in the remarkable relations that Hilbert must have perceived between the structure of Mie's theory of electromagnetism and Einstein's theory of gravitation, as the latter was presented in his 1913/1914 publications and (presumably) also in the Göttingen lectures.

Like Mie's theory, Einstein's *Entwurf* theory was based on a variational principle for a Lagrangian H, here considered to be a function of the gravitational potentials (represented by the components of the metric tensor field $g_{\alpha\beta}$) and their first derivatives. In contrast to Mie, however, Einstein had specified a particular Lagrangian, from which he then derived the gravitational field equations:[31]

---

27 "Ein Atom ist eine Zusammenballung einer größeren Zahl von Elektronen, die durch eine verhältnismäßig dünne Ladung von entgegengesetztem Vorzeichen verkittet sind. Die Atome sind wahrscheinlich von kräftigeren Atmosphären umgeben, die allerdings immer noch so dünn sind, daß sie keine bemerkbaren elektrischen Felder veranlassen, die sich aber vermutlich in den Gravitationswirkungen geltend machen." See (Mie 1912a, 512–513).

28 See (Mie 1913, 5).

29 Compare (Mie 1912a, 534) with (Mie 1913, 29).

30 For notes on a part of Einstein's lectures, see "Nachschrift of Einstein's Wolfskehl Lectures" in (CPAE 6, 586–590). For a discussion of Einstein's Göttingen visit and its possible impact on Hilbert, see (Corry 1999a, 514–517).

31 Our presentation follows Einstein's major review paper, (Einstein 1914b).

$$H = \frac{1}{4}\sum_{\alpha\beta\tau\rho} g^{\alpha\beta}\frac{\partial g_{\tau\rho}}{\partial x_\alpha}\frac{\partial g^{\tau\rho}}{\partial x_\beta}. \tag{6}$$

To be more precise, Einstein was able to derive the empty-space field equations from this Lagrangian. The left-hand side of the gravitational field equations is given by the Lagrangian derivative of (6):[32]

$$\mathfrak{G}_{\mu\nu} = \frac{\partial H\sqrt{-g}}{\partial g^{\mu\nu}} - \sum_\sigma \frac{\partial}{\partial x_\sigma}\left(\frac{\partial H\sqrt{-g}}{\partial g^{\mu\nu}_\sigma}\right) \tag{7}$$

where $g^{\mu\nu}_\sigma \equiv \frac{\partial}{\partial x_\sigma} g^{\mu\nu}$. In the presence of matter, the right-hand side of the field equations is given by the energy-momentum tensor $\mathfrak{T}_{\alpha\beta}$ of matter, so that Einstein's field equations become:

$$\mathfrak{G}_{\sigma\tau} = \kappa\mathfrak{T}_{\sigma\tau}, \tag{8}$$

with the universal gravitational constant κ. In Einstein's *Entwurf* theory, the role of matter as an external source of the gravitational field is not determined by the theory, but rather to be prescribed independently. In the Lagrangian, matter thus appears simply "black-boxed," in the form of a term involving its energy-momentum tensor, rather than as an expression explicitly involving some set of variables describing the constitution of matter:

$$\int\left(\delta H - \kappa\sum_{\mu\nu}\mathfrak{T}_{\mu\nu}\delta g^{\mu\nu}\right)d\tau = 0. \tag{9}$$

Here was a possible point of contact between Mie's and Einstein's theories: Was it possible to conceive of Mie's electromagnetic matter as the source of Einstein's gravitational field? In order to answer this question, evidently one had to study how the energy-momentum tensor $\mathfrak{T}_{\alpha\beta}$ can be derived from terms of Mie's Lagrangian; in particular, what happens if Mie's matter is placed in a four-dimensional spacetime described by an arbitrary metric tensor $g_{\mu\nu}$? This naturally presupposed a reformulation of Mie's theory in generally-covariant form, with an arbitrary metric tensor $g_{\mu\nu}$ replacing the flat one of Minkowski spacetime.

Although most other expressions in his theory are generally-covariant, such as the geodesic equations of motion for a particle in the $g_{\mu\nu}$-field and the expression of energy-momentum conservation in the form of the vanishing covariant divergence of the energy tensor of matter, the field equations of Einstein's 1913/14 theory of gravitation are not. While this lack of general covariance had initially seemed to him to be a blemish on his theory, in late 1913 Einstein convinced himself that he could

32 Magnitudes in Gothic script represent tensor densities with respect to linear transformations.

even demonstrate—by means of the well-known "hole-argument"—that generally-covariant field equations are physically inadmissible because they cannot provide a unique solution for the metric tensor $g_{\mu\nu}$ describing the gravitational field produced by a given matter distribution. The hole argument involves a specific boundary value problem (whether this problem is well posed mathematically is a question that Einstein never considered) for a set of generally-covariant field equations with given sources outside of and boundary values on a "hole" (i.e. a region of spacetime without any sources in it), Einstein showed how to construct infinitely many apparently inequivalent solutions starting from any given solution. From the perspective of the hole argument, as Hilbert realized, if one considers generally-covariant field equations, then in order to pick out a unique solution these equations must be supplemented by four additional non-covariant equations. From the perspective of the 1915 theory of general relativity, however, the hole argument no longer represents an objection against generally-covariant field equations because the class of mathematically distinct solutions generated from an initial solution are not regarded as physically distinct, but merely as different mathematical representations of a single physical situation.[33]

Even in 1913/14 Einstein believed that it might be possible to formulate generally-covariant equations, from which equations (8) would follow by introducing a suitable coordinate restriction.[34] While he actually never found such equations corresponding to (8), he did find four non-covariant coordinate restrictions that he believed characteristic for his theory. He obtained these coordinate restrictions from an analysis of the behavior under coordinate transformations of the variational principle, on which his theory was based. Expressed in terms of the Lagrangian H, these four coordinate restrictions are:

$$B_\mu = \sum_{\alpha\sigma\nu} \frac{\partial^2}{\partial x_\sigma \partial x_\alpha}\left(g^{\nu\alpha}\frac{\partial H\sqrt{-g}}{\partial g^{\mu\nu}_\sigma}\right) = 0. \tag{10}$$

Einstein regarded these restrictions as making evident the non-general covariance of his theory; indeed he believed them just restrictive enough to avoid the hole-argument. Einstein also required the existence of a gravitational energy-momentum complex (non-tensorial) guaranteeing validity of four energy-momentum conservation equations for the combined matter and gravitational fields. His theory thus involved 10 field equations, 4 coordinate restrictions, and 4 conservation equations — in all 18 equations for the 10 gravitational potentials $g_{\mu\nu}$.

Einstein used the consistency of this overdetermined system as a criterion for the choice of a Lagrangian, imposing the condition that the field equations together with the energy-momentum conservation equations should yield the coordinate restric-

33 See (Stachel 1989; 71–81, sections 3 and 4).

34 See, e.g., (Einstein 1914a, 177–178). It is unclear whether Einstein expected the unknown generally-covariant equations to be of higher order than second.

tions (10). For this purpose, he assumed a general Lagrangian H depending on $g_{\mu\nu}$ and $g_{\mu\nu,\kappa}$, and then examined the four equations implied by the assumption of energy-momentum conservation for the field equations resulting from this Lagrangian. Formulating energy-momentum conservation as the requirement that the covariant divergence of the energy-momentum tensor density $\mathfrak{T}_\sigma^{\ \nu}$ has to vanish, and using the field equations (8), he first obtained:

$$\begin{aligned} \nabla_\nu \mathfrak{T}_\sigma^{\ \nu} \equiv \sum_{\nu\tau} \frac{\partial}{\partial x_\nu}(g^{\tau\nu}\mathfrak{T}_{\sigma\tau}) + \frac{1}{2}\sum_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}\mathfrak{T}_{\mu\nu} = 0 \Rightarrow \\ \sum_{\nu\tau} \frac{\partial}{\partial x_\nu}(g^{\tau\nu}\mathfrak{E}_{\sigma\tau}) + \frac{1}{2}\sum_{\mu\nu}\frac{\partial g^{\mu\nu}}{\partial x_\sigma}\mathfrak{E}_{\mu\nu} = 0, \end{aligned} \tag{11}$$

and then:

$$\sum_\nu \frac{\partial S_\sigma^\nu}{\partial x_\nu} - B_\sigma = 0, \tag{12}$$

with B_σ given by (10) and:

$$S_\sigma^\nu = \sum_{\mu\tau}\left(g^{\nu\tau}\frac{\partial H\sqrt{-g}}{\partial g^{\sigma\tau}} + g_\mu^{\nu\tau}\frac{\partial H\sqrt{-g}}{\partial g_\mu^{\sigma\tau}} + \frac{1}{2}\partial_\sigma^\nu H\sqrt{-g} - \frac{1}{2}g_\sigma^{\mu\tau}\frac{\partial H\sqrt{-g}}{\partial g_\nu^{\mu\tau}}\right). \tag{13}$$

By requiring that:

$$S_\sigma^\nu \equiv 0, \tag{14}$$

an equation that indeed is satisfied for the Lagrangian (6), it follows that (12) entails no new conditions beyond (10). In other words, for the "right" Lagrangian, the coordinate restrictions required by the hole-argument follow from energy-momentum conservation. In late 1915 Einstein found that his argument for the uniqueness of the Lagrangian, and thus for the uniqueness of the field equations, is fallacious;[35] and this insight helped to motivate him to return to generally-covariant field equations.

If one disregards the wealth of successful predictions of Newtonian gravitation theory that also buttressed Einstein's theory of 1913/14, that theory might appear almost as speculative as Mie's theory of matter. On the one hand, Einstein had been able to make several predictions based on his theory, such as the perihelion shift of Mercury, the deflection of light in a gravitational field, and gravitational redshift, that, at least in principle, could be empirically checked. On the other hand, none of these conclusions had actually received such support by the time Hilbert turned to Einstein's work: indeed, the calculated perihelion shift was in disaccord with observation.

35 For a historical discussion, see (Norton 1984) and "Untying the Knot …" (in vol. 2 of this series).

2.3 Hilbert's Research Program

To a mathematician of Hilbert's competence, Einstein's 1913/1914 theory must have appeared somewhat clumsy. In particular, it left several specifically mathematical questions open, such as the putative existence of the corresponding generally-covariant equations mentioned above; how the field equations (8) result from these generally-covariant equations by means of the coordinate restrictions (10); whether the hole argument for generally-covariant equations is better applied to boundary values on an open space-like hypersurface (the Cauchy problem) or a closed hypersurface (Einstein's formulation); and the closely–related question of the number of independent equations for the gravitational potentials in Einstein's system. Such questions presumably suggested to Hilbert a rather well-circumscribed research program that, taken together with his interest in Mie's theory of matter, amounted to the search for an "axiomatic synthesis" of the two speculative physical theories.

In consequence, Hilbert's initial program presumably comprised:[36]

1. a generally-covariant reformulation of both Mie's and Einstein's theories with the intention of deriving both from a single variational principle for a Lagrangian that depends on both Mie's electrodynamical and Einstein's gravitational variables;
2. an examination of the possibility of replacing Einstein's unspecified energy-momentum tensor for matter by one following from Mie's Lagrangian;
3. a further examination of the non-uniqueness of solutions to generally-covariant equations, involving a study of the question of the number of independent equations, and finally
4. the identification of coordinate restrictions appropriate to delimit a unique solution and an examination of their relation to energy-momentum conservation.

Even prior to looking at Hilbert's attempt to realize such a synthesis of Mie's and Einstein's approaches, it is clear that such a program would fit perfectly into Hilbert's axiomatic approach to physics. Indeed, the realization of this suggested initial program would: constitute a clarification of the logical and mathematical foundations of already existing physical theories in their own terms; represent the synthesis of different theories by combination of logically independent elements within one and the same formalism (in this case incorporation of Mie's variables and Einstein's variables in the same Lagrangian); replace the unspecified character of the material sources entering Einstein's theory with a daring theory of their electromagnetic nature, formulated in mathematical terms, thus shifting the boundary between experience and mathematical deduction in favor of the latter.

Unfortunately, there is no direct evidence that Hilbert developed and pursued some such research program in the course of his work in the second half of 1915 on Mie's and Einstein's theories. We have no "Göttingen notebook" that would be equivalent to Einstein's "Zurich Notebook," documenting in detail the heuristics that Hil-

36 For a similar attempt to reconstruct Hilbert's research program, see (Sauer 1999, 557–559).

bert followed.[37] However, now we have the first proofs of Hilbert's first communication that (as we have argued)[38] provide a glimpse into his thinking prior to his assimilation of Einstein's definitive paper on general relativity. In the next section we shall argue that the proofs version of Hilbert's theory can be interpreted as the result of pursuing just such a research program as that sketched above.

3. HILBERT'S ATTEMPT AT A THEORY OF EVERYTHING: THE PROOFS OF HIS FIRST PAPER

In this section we shall attempt to reconstruct Hilbert's heuristics from the Proofs and published versions of his first paper (Hilbert 1916), hereafter, Proofs and Paper 1. We will begin by reconstructing from the Proofs and other contemporary documents, the first step in the realization of Hilbert's program. This crucial step, an attempt to explore the first two points of the program, was the establishment of a relation between Mie's energy-momentum tensor and the variational derivative with respect to the metric of Mie's Lagrangian.[39] Next, we attempt to reconstruct Hilbert's calculation of Mie's energy-momentum tensor from the Born-Mie Lagrangian. We then examine the consequences of this derivation for the concept of energy, and thus for the further exploration of the second point of his program. We then discuss how these results suggest a new perspective on the relation between Mie's and Einstein's theories, from which gravitation appears more fundamental than electrodynamics. Seen from this perspective, the third point of Hilbert's program, the question of uniqueness of solutions to generally-covariant equations, took on a new significance: Hilbert turned Einstein's argument that only a non-covariant theory can make physical sense into an instrument for the synthesis of electromagnetism and gravitation. Coming to the fourth point of Hilbert's program, we show how he united his energy concept with the requirement of restricting general covariance. Finally, after examining Hilbert's attempt to derive the electromagnetic field equations from the gravitational ones, we discuss Hilbert's rearrangement of his results in the form of an axiomatically constructed theory, which he presented in the Proofs of Paper 1.

3.1 The First Result

At some point in late summer or fall of 1915, Hilbert must have discovered a relation between the energy-momentum tensor following from Mie's theory of matter, the Born-Mie Lagrangian L, and the metric tensor representing the gravitational poten-

37 Einstein's search for gravitational field equations in the winter of 1912/13 is documented in the so-called Zurich Notebook, partially published as Doc. 10 of (CPAE 4). Einstein's research project has been reconstructed in volumes 1 and 2 of this series. See, in particular, "Pathways out of Classical Physics …" (in vol. 1 of this series).

38 In (Corry, Renn, and Stachel 1997).

39 Henceforth, mention of the variational derivative of a Lagrangian, without further indication, always means with respect to the metric tensor.

tials in Einstein's theory of gravitation. In the Proofs and the published version of Paper 1, as well as in his contemporary correspondence, Hilbert emphasized the significance of this discovery for his understanding of the relation between Mie's and Einstein's theories. In the Proofs he wrote:

> Mie's electromagnetic energy tensor is nothing but the generally invariant tensor that results from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in the limit (25) [i.e. the equation $g_{\mu\nu} = \delta_{\mu\nu}$] — a circumstance that gave me the first hint of the necessary close connection between Einstein's general relativity theory and Mie's electrodynamics, and which convinced me of the correctness of the theory here developed.[40]

Hilbert expressed himself similarly in a letter of 13 November 1915 to Einstein:

> I derived most pleasure in the discovery, already discussed with Sommerfeld, that the usual electrical energy results when a certain absolute invariant is differentiated with respect to the gravitation potentials and then g is set = 0,1.[41]

On the basis of our suggested reconstruction of Hilbert's research program, it is possible to suggest what might have led him to this relation. We assume that he attempted to realize the first two steps, that is to reformulate Mie's Lagrangian in a generally-covariant setting and replace the energy-momentum tensor term in Einstein's variational principle by a term corresponding to Mie's theory. Considering (9), this would imply an expression such as $\delta H + \delta L$ under the integral, where H corresponds to Einstein's original Lagrangian and L to a generally-covariant form of Mie's Lagrangian. If the variation of Mie's Lagrangian is regarded as representing the energy-momentum tensor term, one obtains:

$$\delta L = -\kappa \sum_{\mu\nu} \mathfrak{T}_{\mu\nu} \delta g^{\mu\nu}, \tag{15}$$

where $\mathfrak{T}_{\mu\nu}$ should now be the energy-momentum tensor of Mie's theory. It may well have been an equation of this form, following from the attempt to replace the unspecified source-term in Einstein's field equations by a term depending on the generally-covariant form of Mie's Lagrangian, that first suggested to Hilbert that the energy-momentum tensor of Mie's theory could be the variational derivative of Mie's Lagrangian.

---

40 "der Mie'sche elektromagnetische Energietensor ist also nichts anderes als der durch Differentiation der Invariante L nach den Gravitationspotentialen $g^{\mu\nu}$ entstehende allgemein invariante Tensor beim Übergang zum Grenzfall (25) [i.e. the equation $g_{\mu\nu} = \delta_{\mu\nu}$] — ein Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen und mir die Überzeugung von der Richtigkeit der hier entwickelten Theorie gegeben hat." (Proofs, 10)

41 "Hauptvergnügen war für mich die schon mit Sommerfeld besprochene Entdeckung, dass die gewöhnliche elektrische Energie herauskommt, wenn man eine gewisse absolute Invariante mit den Gravitationspotentialen differenziert und [d]ann $g = 0, 1$ setzt." David Hilbert to Einstein, 13 November 1915, (CPAE 8, 195). Unless otherwise noted, all translations are based on those in the companion volumes to the Einstein edition, but often modified.

If he followed the program outlined above, Hilbert would have assumed that the Lagrangian has the form:

$$H = K + L, \tag{16}$$

where K represents the gravitational part and L the electromagnetic. Indeed, this form of the Lagrangian is used both in the Proofs and the published version of Paper 1.[42]

In Paper 1, Hilbert derived a relation of the form:

$$-2\sum_{\mu}\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}g^{\mu m} = T_{\nu}^{m}, \tag{17}$$

where T_{ν}^{m} stands for the energy-momentum tensor density of Mie's theory.[43] This relation, which is exactly what one would expect on the basis of (15), could have suggested to Hilbert that a deep connection must exist between the nature of spacetime as represented by the metric tensor and the structure of matter as represented by Mie's theory.

3.2 Mie's Energy-Momentum Tensor as a Consequence of Generally-Covariant Field Equations

The strategy Hilbert followed to derive (17) can be reconstructed from the two versions of his paper. It consisted in following as closely as possible the standard variational techniques applied, for instance, to derive Lagrange's equations from a variational principle.[44] In Hilbert's paper, a similar variational problem forms the core of his theory. He describes his basic assumptions in two axioms:[45]

> Axiom I (Mie's axiom of the world function): *The law governing physical processes is determined through a world function H that contains the following arguments*:
>
> $$g_{\mu\nu},\quad g_{\mu\nu l}=\frac{\partial g_{\mu\nu}}{\partial w_l},\qquad g_{\mu\nu lk}=\frac{\partial^2 g_{\mu\nu}}{\partial w_l \partial w_k},$$
> $$q_s,\quad q_{sl}=\frac{\partial q_s}{\partial w_l}\qquad\qquad (l,k=1,2,3,4), \tag{18}$$
>
> *where the variation of the integral*

42 In the Proofs it was presumably introduced on the upper part of p. 8, which unfortunately is cut off.

43 See (Proofs, 10; Hilbert 1916, 404). Note that Hilbert uses an imaginary fourth coordinate, so that the minus sign emerges automatically in the determinant of the metric; he does not explicitly introduce the energy-momentum tensor T_{ν}^{m}.

44 See, for example, (Caratheodory 1935).

45 See also (Hilbert 1916, 396).

$$\int H\sqrt{g}\,d\tau \qquad (g = |g_{\mu\nu}|, \quad d\tau = dw_1 dw_2 dw_3 dw_4) \tag{19}$$

must vanish for each of the fourteen potentials $g_{\mu\nu}, q_s$.[46]

[The w_s are Hilbert's notation for an arbitrary system of coordinates.]

> Axiom II (axiom of general invariance): *The world function* H *is invariant with respect to an arbitrary transformation of the world parameters* w_s.[47]

Starting from an arbitrary invariant J, Hilbert formed a differential expression from it depending on $g^{\mu\nu}, g_l^{\mu\nu}, g_{lk}^{\mu\nu}, q_s, q_{sk}$, which in the published version of his paper he called PJ. He defined the operator P as follows:[48]

$$\begin{aligned} P &= P_g + P_q, \\ P_g &= \sum_{\mu,\nu,l,k} \left(p^{\mu\nu}\frac{\partial}{\partial g^{\mu\nu}} + p_l^{\mu\nu}\frac{\partial}{\partial g_l^{\mu\nu}} + p_{lk}^{\mu\nu}\frac{\partial}{\partial g_{lk}^{\mu\nu}} \right), \\ P_q &= \sum_{l,k} \left(p_l\frac{\partial}{\partial q_l} + p_{lk}\frac{\partial}{\partial q_{lk}} \right), \end{aligned} \tag{20}$$

where $p^{\mu\nu}$ and p_l are arbitrary variations of the metric tensor and the electromagnetic four-potentials, respectively. Thus:

$$PJ = \sum_{\mu,\nu,l,k} \left(p^{\mu\nu}\frac{\partial J}{\partial g^{\mu\nu}} + p_l^{\mu\nu}\frac{\partial J}{\partial g_l^{\mu\nu}} + p_{lk}^{\mu\nu}\frac{\partial J}{\partial g_{lk}^{\mu\nu}} + p_l\frac{\partial J}{\partial q_l} + p_{lk}\frac{\partial J}{\partial q_{lk}} \right). \tag{21}$$

In the mathematical terminology of the time, PJ is a "polarization" of J.[49]

As we shall see, it is possible to derive from PJ identities that realize Hilbert's goal, the derivation of (17). His procedure is described more explicitly in the published version of Paper 1, and since we assume that on this point there was no significant development of Hilbert's thinking after the Proofs, our reconstruction will make use of the published version.

In modern terminology, if $p^{\mu\nu}$ and p_l are those special variations generated by dragging the metric and the electromagnetic potentials over the manifold with some vector field p^s; i.e., if they are the Lie derivatives of the metric and the electromagnetic potentials with respect to p^s,[50] then PJ must be the Lie derivative of J with

---

46 "Axiom I (Mie's Axiom von der Weltfunktion): *Das Gesetz des physikalischen Geschehens bestimmt sich durch eine Weltfunktion H, die folgende Argumente enthält*: [(18); (1) and (2) in the original text] *und zwar muß die Variation des Integrals* [(19)] *für jedes der 14 Potentiale* $g_{\mu\nu}, q_s$ verschwinden." (Proofs, 2) The q_s are the electromagnetic four potentials.

47 "Axiom II (Axiom von der allgemeinen Invarianz): *Die Weltfunktion H ist eine Invariante gegenüber einer beliebigen Transformation der Weltparameter* w_s." (Proofs, 2)

48 See (Hilbert 1916, 398–399). Compare (Proofs, 4 and 7).

49 See, e.g., (Kerschensteiner 1887, §2).

respect to p^s. On the other hand, since J is a scalar invariant, the Lie derivative of this scalar with respect to p^s can be written directly, so that:

$$\sum_s \frac{\partial J}{\partial w_s} p^s = PJ. \tag{22}$$

With a little work,[51] equation (22) can be rewritten in the form of equation (23) below. This is the content of Hilbert's Theorem II, both in the Proofs and in Paper 1:

> Theorem II. If J is an invariant depending on $g^{\mu\nu}$, $g^{\mu\nu}_l$, $g^{\mu\nu}_{lk}$, q_s, q_{sk}, then the following is always identically true in all its arguments and for every arbitrary contravariant vector p^s:

$$\begin{aligned}\sum_{\mu,\nu,l,k}\left(\frac{\partial J}{\partial g^{\mu\nu}}\Delta g^{\mu\nu} + \frac{\partial J}{\partial g^{\mu\nu}_l}\Delta g^{\mu\nu}_l + \frac{\partial J}{\partial g^{\mu\nu}_{lk}}\Delta g^{\mu\nu}_{lk}\right)\\ + \sum_{s,k}\left(\frac{\partial J}{\partial q_s}\Delta q_s + \frac{\partial J}{\partial q_{sk}}\Delta q_{sk}\right) = 0;\end{aligned} \tag{23}$$

where

$$\begin{aligned}\Delta g^{\mu\nu} &= \sum_m (g^{\mu m} p^\nu_m + g^{\nu m} p^\mu_m),\\ \Delta g^{\mu\nu}_l &= -\sum_m g^{\mu\nu}_m p^m_l + \frac{\partial \Delta g^{\mu\nu}}{\partial w_l},\\ \Delta g^{\mu\nu}_{lk} &= -\sum_m (g^{\mu\nu}_m p^m_{lk} + g^{\mu\nu}_{lm} p^m_k + g^{\mu\nu}_{km} p^m_l) + \frac{\partial^2 \Delta g^{\mu\nu}}{\partial w_l \partial w_k},\\ \Delta q_s &= -\sum_m q_m p^m_s,\\ \Delta q_{sk} &= -\sum_m q_{sm} p^m_k + \frac{\partial \Delta q_s}{\partial w_k}.\end{aligned} \tag{24}$$

[52]

Hilbert next applies Theorem II to the electromagnetic part L of his Lagrangian $H = K + L$, with the assumption that L only depends on the metric $g^{\mu\nu}$, the elec-

---

50 Here $p^{\mu\nu}$ corresponds, in modern terms, to the Lie derivative of the contravariant form of the metric tensor with respect to the arbitrary vector p^j. Hilbert writes:

$$p^{\mu\nu} = \sum_s (g^{\mu\nu}_s p^s - g^{\mu s} p^\nu_s - g^{\nu s} p^\mu_s), \quad \left(p^j_s = \frac{\partial p^j}{\partial w_s}\right),$$

and similarly for the Lie derivatives of the electromagnetic potentials. While the term "Lie derivative" was only introduced in 1933 by W. Slebodzinski (see Slebodzinski 1931), it was well known in Hilbert's time that the basic idea came from Lie; see for example (Klein 1917, 471): "For this purpose one naturally determines, as Lie in particular has done in his numerous relevant publications, the formal changes that result from an arbitrary infinitesimal transformation." ("Zu diesem Zwecke bestimmt man natürlich, wie dies insbesondere Lie in seinen zahlreichen einschlägigen Veröffentlichungen getan hat, die formellen Änderungen, welche sich bei einer beliebigen infinitesimalen Transformation ... ergeben ... .") According to Schouten, the name "Lie differential" was proposed by D. Van Dantzig; see (Schouten and Struik 1935, 142).

tromagnetic potentials q_s and their derivatives q_{sk}, but *not* on the derivatives of the metric tensor. This gives the identity:[53]

$$\sum_{\mu,\nu,m} \frac{\partial L}{\partial g^{\mu\nu}}(g^{\mu m}p^{\nu}_{m} + g^{\nu m}p^{\mu}_{m}) - \sum_{s,m} \frac{\partial L}{\partial q_s} q_m p^m_s - \sum_{s,k,m} \frac{\partial L}{\partial q_{sk}}(q_{sm}p^m_k + q_{mk}p^m_s + q_m p^m_{sk}) = 0. \tag{25}$$

Since the vector field p^s is arbitrary, its coefficients as well as the coefficients of its first and second derivatives must vanish identically. Hilbert drew two conclusions, which he interpreted as strong links between a generally-covariant variational principle and Mie's theory of matter. The first concerns the form in which the electromag-

51 See (Proofs, 7–8; Hilbert 1916, 398). The equivalence of (22) and (23) is shown as follows: Since J depends on w_s through $g^{\mu\nu}$, $g^{\mu\nu}_m$, $g^{\mu\nu}_{mk}$, q_m and q_{mk} it follows that:

$$\frac{\partial J}{\partial w_s} = \frac{\partial J}{\partial g^{\mu\nu}} \cdot g^{\mu\nu}_s + \frac{\partial J}{\partial g^{\mu\nu}_m} \cdot g^{\mu\nu}_{sm} + \frac{\partial J}{\partial g^{\mu\nu}_{mk}} \cdot g^{\mu\nu}_{smk} + \frac{\partial J}{\partial q_m} \cdot q_{ms} + \frac{\partial J}{\partial q_{mk}} \cdot q_{mks}.$$

On the other hand, PJ is the Lie derivative of J through its dependence on $g^{\mu\nu}$, $g^{\mu\nu}_m$ q_m and q_{mk}, so:

$$PJ = \frac{\partial J}{\partial g^{\mu\nu}} \cdot p^{\mu\nu} + \frac{\partial J}{\partial g^{\mu\nu}_m} \cdot p^{\mu\nu}_m + \frac{\partial J}{\partial g^{\mu\nu}_{mk}} \cdot p^{\mu\nu}_{mk} + \frac{\partial J}{\partial q_m} \cdot p_m + \frac{\partial J}{\partial q_{mk}} \cdot p_{mk}$$

where $p^{\mu\nu}$ $p^{\mu\nu}_m$, $p^{\mu\nu}_{mk}$, p_m and p_{mk} stand for the Lie derivatives with respect to the vector field p^k of $g^{\mu\nu}$, $g^{\mu\nu}_m$, $g^{\mu\nu}_{mk}$, q_m and q_{mk} respectively (Hilbert's notation). Rewriting (24) in terms of the definition of the Lie derivatives of $g^{\mu\nu}$, $g^{\mu\nu}_m$, $g^{\mu\nu}_{mk}$, q_m and q_{mk}, we easily get:

$$\begin{aligned}
\Delta g^{\mu\nu} &= \sum_m g^{\mu\nu}_m p^m - p^{\mu\nu},\\
\Delta g^{\mu\nu}_l &= \sum_m g^{\mu\nu}_{ml} p^m - p^{\mu\nu}_l,\\
\Delta g^{\mu\nu}_{lk} &= \sum_m g^{\mu\nu}_{mlk} p^m - p^{\mu\nu}_{lk},\\
\Delta q_s &= \sum_m q_{sm} p^m - p_s,\\
\Delta q_{sk} &= \sum_m q_{smk} p^m - p_{sk}.
\end{aligned}$$

Inserting these expressions into (23), and using the equations for $\frac{\partial J}{\partial w_s}$ and PJ at the beginning of this note, one sees that (23) reduces to:

$$\frac{\partial J}{\partial w_s} \cdot p^s - PJ = 0,$$

which is equivalent to (22).

netic potentials enter the Lagrangian, the second concerns the relation between this Lagrangian and Mie's energy-momentum tensor.

From Hilbert's requirements on L —that it be a generally-invariant scalar that does not depend on the derivatives of the metric tensor—he was able to show that the derivatives of the electromagnetic potentials can only enter it in the form characteristic of Mie's theory (see (5)). Setting the coefficients of p^m_{sk} in (25) equal to zero, and remembering that $p^m_{sk} = p^m_{ks}$, one obtains:

$$\left(\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}}\right) q_m = 0. \tag{26}$$

Since q_m cannot vanish identically, it follows that:

$$\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0, \tag{27}$$

which mean that the q_{ik} only enter L in the antisymmetric combination familiar from Mie's theory:

$$M_{ks} = q_{sk} - q_{ks}. \tag{28}$$

Thus, apart from the potentials themselves, L depends only on the components of the tensor M:

$$M = \mathrm{Rot}(q_s), \tag{29}$$

the familiar electromagnetic "six vector." Hilbert emphasized:

> *This result here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*[54]

In order to explicitly establish the relation between his theory and Mie's, Hilbert points out that L must be a function of four invariants.[55] Hilbert only gave what he considered to be the "two simplest" of the generally-covariant generalizations of these invariants:

$$Q = \sum_{k,\,l,\,m,\,n} M_{mn} M_{lk} g^{mk} g^{nl} \tag{30}$$

---

52 "Theorem II. Wenn J eine von $g^{\mu\nu}$, $g^{\mu\nu}_l$, $g^{\mu\nu}_{lk}$, q_s, q_{sk} abhängige Invariante ist, so gilt stets identisch in allen Argumenten und für jeden willkürlichen kontravarianten Vektor p^s [(23)] dabei ist: [(24)]."

53 See (Proofs, 9; Hilbert 1916, 403).

54 "*Dieses Resultat ergibt sich hier wesentlich als Folge der allgemeinen Invarianz, also auf Grund von Axiom II.*" (Proofs, 10) In the published version this passage reads: "*This result, which determines the character of Maxwell's equations in the first place, here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*" ("*Dieses Resultat, durch welches erst der Charakter der Maxwellschen Gleichungen bedingt ist, ergibt sich hier wesentlich als Folge der allgemeinen Invarianz, also auf Grund von Axiom II.*") See (Hilbert 1916, 403).

55 See (Proofs, 13, and Hilbert 1916, 407). Here Hilbert followed the papers of Mie and Born; see, in particular, (Born 1914).

and:

$$q = \sum_{k,\,l} q_k q_l g^{kl}. \tag{31}$$

According to Hilbert, the simplest expression that can be formed by analogy to the gravitational part of the Lagrangian K is:[56]

$$L = \alpha Q + f(q), \tag{32}$$

where $f(q)$ is any function of q and α a constant. In order to recover Mie's main example (see (1)) from this more general result, Hilbert considers the following specific functional dependence:

$$L = \alpha Q + \beta q^3, \tag{33}$$

which corresponds to the Lagrangian given by Mie. In contrast to Mie, Hilbert does not even allude to the physical problems associated with this Lagrangian. And in contrast to Einstein, at no point does Hilbert introduce the Newtonian coupling constant into his equations, so that his treatment of gravitation remains as "formalistic" as that of electromagnetism.

The second consequence Hilbert drew from (25), which corresponds to what we have called above "Hilbert's first results" (see (17)), concerns Mie's energy-momentum tensor. Setting the coefficient of p_m^{ν} equal to zero and using (27), he obtained:[57]

$$2\sum_{\mu} \frac{\partial L}{\partial g^{\mu\nu}} g^{\mu m} - \frac{\partial L}{\partial q_m} q_{\nu} - \sum_{s} \frac{\partial L}{\partial M_{ms}} M_{\nu s} = 0, \quad (\mu = 1, 2, 3, 4). \tag{34}$$

Noting that:

$$2\sum_{\mu} \frac{\partial L}{\partial g^{\mu\nu}} g^{\mu m} = \frac{2}{\sqrt{g}} \cdot \sum_{\mu} \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} + L \cdot \delta_{\nu}^{\ m}, \tag{35}$$

(34) can be rewritten:

$$-2\sum_{\mu} \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} = \sqrt{g}\left\{ L\delta_{\nu}^{m} - \frac{\partial L}{\partial q_m} q_{\nu} - \sum_{s} \frac{\partial L}{\partial M_{ms}} M_{\nu s} \right\}, \tag{36}$$

$$(\mu = 1, 2, 3, 4) \quad (\delta_{\nu}^{\mu} = 0, \ \mu \neq \nu, \ \delta_{\mu}^{\mu} = 1).$$

The right-hand side of this equation is the generally-covariant generalization of Mie's energy-momentum tensor. It is this equation that inspired Hilbert's remark about the

56 Note that Q is the term that gives rise to Maxwell's equations and that q cannot be used if the resulting theory is to be gauge invariant. See (Born and Infeld 1934).

57 See (Proofs, 10; Hilbert 1916, 404).

"Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen ... hat", quoted above (p. 873). Hilbert had shown that characteristic properties of Mie's Lagrangian follow from its generally-covariant generalization, a result he interpreted as indicating that gravitation must be conceived as being more fundamental than electromagnetism, as his later work indicates.

3.3 The Definition of Energy

While (36) shows a strong link between a generally-covariant L and Mie's energy momentum tensor, it does not answer the question of how energy-momentum conservation is to be conceived in Hilbert's theory. Hilbert's theory does not allow the interpretation of an energy-momentum tensor for matter as an external source, as does that of Einstein; so Hilbert could not start from a conservation law for matter in Minkowski spacetime and simply generalize it to the case in which a gravitational field is present. Such a procedure would have conflicted with Hilbert's heuristic, according to which matter itself is conceived in terms of electromagnetic fields that, in turn, arise in conjunction with, or even as an effect of, gravitational fields.

Hilbert's heuristic for finding an appropriate definition of energy seems to be governed by a formal criterion related to his understanding of energy conservation in classical physics, as well as by a criterion with a more specific physical meaning related to the results he expected from Mie's theory. Hilbert's formal criterion is well described in a passage in his summer-semester 1916 lectures on the foundations of physics, a passage which occurs in a discussion of energy-momentum conservation in Mie's theory:

> The energy concept comes from just writing Lagrange's equations in the form of a divergence, and defining as energy what is represented as divergent.[58]

As for Hilbert's physical criterion, any definition of the energy must be compatible with his insight that the variational derivative of Mie's Lagrangian yields the electromagnetic energy-momentum tensor.

Hilbert's treatment of energy conservation in the Proofs and in Paper 1 is not easy to follow. This difficulty was felt by Hilbert's contemporaries; both Einstein and Klein had their problems with it.[59] Nevertheless, as will become clear in what follows, Hilbert's discussion was guided by the heuristic criteria mentioned above. He proceeded in three steps:

- he first identified an energy expression consisting of a sum of divergence terms (Satz 1 in the Proofs):

---

58 "Der Energiebegriff kommt eben daher, dass man die Lagrangeschen Gleichungen in Divergenzform schreibt, und das, was unter der Divergenz steht, als Energie definiert." Die Grundlagen der Physik I, Ms. Vorlesung SS 1916, 98 (D. Hilbert, Bibliothek des Mathematischen Seminars, Universität Göttingen); from here on "SS 1916 Lectures."

- he then formulated a divergence equation for his energy expression in analogy to classical and special-relativistic results (Satz 2 in the Proofs), and imposed this equation as a requirement implying coordinate restrictions (Axiom III):
- finally, he showed that his energy expression can be related to Mie's energy-momentum tensor (the real justification of his choice).

Here we focus on the first and last of these points, deferring the issue of coordinate restrictions to a subsequent section ("Energy-momentum conservation and coordinate restrictions").

As in his derivation of the connection between Mie's energy-momentum tensor and the variational derivative of the Lagrangian, Hilbert's starting point was his generally-covariant variational principle. However, he now proceeded somewhat differently. Instead of focussing on the electromagnetic part L, he considered the entire Lagrangian H, but now neglected the derivatives with respect to the electromagnetic potentials, i.e. the contribution of the term P_q to P (see (20)). Accordingly, Hilbert forms the expression:[60]

$$J^{(p)} = \sum_{\mu,\nu} \frac{\partial H}{\partial g^{\mu\nu}} p^{\mu\nu} + \sum_{\mu,\nu,k} \frac{\partial H}{\partial g^{\mu\nu}_k} p^{\mu\nu}_k + \sum_{\mu,\nu,k,l} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} p^{\mu\nu}_{kl}, \tag{37}$$

where $p^{\mu\nu}$ corresponds, as we have seen, to the Lie derivative of the metric tensor with respect to the arbitrary vector p^j. By partial integration, Hilbert transforms this expression into:

$$\sqrt{g} J^{(p)} = -\sum_{\mu,\nu} H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} p^{\mu\nu} + E + D^{(p)}, \tag{38}$$

with:

---

59 In (Klein 1917, 475), Klein quotes from a letter he had written to Hilbert concerning the latter's energy expression in Paper 1: "But I find your equations so complicated that I have not attempted to redo your calculations." ("Ich finde aber Ihre Formeln so kompliziert, daß ich die Nachrechnung nicht unternommen habe.") In a letter, in which Einstein asked Hilbert for a clarification of the latter's energy theorem, he wrote: "Why do you make it so hard for poor mortals by withholding the technique behind your ideas? It surely does not suffice for the thoughtful reader if, although able to verify the correctness of your equations, he cannot get a clear view of the overall plan of the analysis." ("Warum machen Sie es dem armen Sterblichen so schwer, indem Sie ihm die Technik Ihres Denkens vorenthalten? Es genügt doch dem denkenden Leser nicht, wenn er zwar die Richtigkeit Ihrer Gleichungen verifizieren aber den Plan der ganzen Untersuchung nicht überschauen kann.") See Einstein to David Hilbert, 30 May 1916, (CPAE 8, 293). In a letter to Paul Ehrenfest, Einstein expressed himself even more drastically with respect to what he perceived as the obscurity of Hilbert's heuristic: "Hilbert's description doesn't appeal to me. It is unnecessarily specialized as concerns "matter," unnecessarily complicated, and not above-board (=Gauss-like) in structure (feigning the super-human through camouflaging the methods)." ("Hilbert's Darstellung gefällt mir nicht. Sie ist unnötig speziell, was die 'Materie' anbelangt, unnötig kompliziert, nicht ehrlich (=Gaussisch) im Aufbau (Vorspiegelung des Übermenschen durch Verschleierung der Methoden).") See Einstein to Paul Ehrenfest, 24 May 1916, (CPAE 8, 288).

60 See (Proofs, 5ff.).

$$
\begin{aligned}
E \;=\; & \sum_{\mu,\,\nu,\,s,\,k,\,l}\Big(H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}g^{\mu\nu}_{s}+\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}}g^{\mu\nu}_{s}+\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{k}}g^{\mu\nu}_{sk}+\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{kl}}g^{\mu\nu}_{skl}\Big)p^{s}\\
& -\sum_{\mu,\,\nu,\,s}(g^{\mu s}p^{\nu}_{s}+g^{\nu s}p^{\mu}_{s})[\sqrt{g}H]_{\mu\nu} \qquad (39)\\
& +\sum_{\mu,\,\nu,\,s,\,k,\,l}\Big(\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{k}}g^{\mu\nu}_{s}+\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}}g^{\mu\nu}_{sl}-g^{\mu\nu}_{s}\frac{\partial}{\partial w_{l}}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}}\Big)p^{s}_{k},
\end{aligned}
$$

and:

$$
\begin{aligned}
D^{(p)} \;=\; & \sum_{\mu,\,\nu,\,s,\,k,\,l}\Big\{-\frac{\partial}{\partial w_{k}}\Big(\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{k}}(g^{\mu s}p^{\nu}_{s}+g^{\nu s}p^{\mu}_{s})\Big)\\
& +\frac{\partial}{\partial w_{k}}\Big((p^{\nu}_{s}g^{\nu s}+p^{\mu}_{s}g^{\nu s})\frac{\partial}{\partial w_{l}}\Big(\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{kl}}\Big)\Big) \qquad (40)\\
& +\frac{\partial}{\partial w_{l}}\Big(\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{kl}}\Big(\frac{\partial p^{\mu\nu}}{\partial w_{k}}-g^{\mu\nu}_{sk}p^{s}\Big)\Big)\Big\}.
\end{aligned}
$$

Hilbert had thus succeeded in splitting off a divergence term $D^{(p)}$ from the original expression $J^{(p)}$. By integrating over some region, $D^{(p)}$ could be converted into a surface term, and thus eliminated by demanding that p^{s} and its derivatives vanish on the boundary of that region.[61] So it would be possible to extract an energy expression from the remainder of $J^{(p)}$ if a way could be found to deal with the first term $\sum_{\mu,\,\nu}H\frac{\partial}{\partial g^{\mu\nu}}\sqrt{g}p^{\mu\nu}$.

Ultimately, the justification for choosing E as the energy expression depends, of course, on the possibility of a physical interpretation of this expression. As we shall see, for Hilbert this meant an interpretation in terms of Mie's theory. But, first of all, he had to show that E can be represented as a sum of divergences. For this purpose, Hilbert introduced yet another decomposition of $J^{(p)}$, derived from a generalization of (37). As we have indicated earlier, this equation may be identified as a special case of a "polarization" of the Lagrangian H with respect to the contravariant form of the metric $g^{\mu\nu}$: If one takes an arbitrary contravariant tensor $h^{\mu\nu}$, one obtains for the "first polar" of H:

$$
J^{(h)} \;=\; \sum_{\mu,\,\nu}\frac{\partial H}{\partial g^{\mu\nu}}h^{\mu\nu}+\sum_{\mu,\,\nu,\,k}\frac{\partial H}{\partial g^{\mu\nu}_{k}}h^{\mu\nu}_{k}+\sum_{\mu,\,\nu,\,k,\,l}\frac{\partial H}{\partial g^{\mu\nu}_{kl}}h^{\mu\nu}_{kl}. \qquad (41)
$$

Applying integration by parts to this expression, Hilbert obtained:

61 Die Grundlagen der Physik II, Ms. Vorlesung WS 1916/17, 186 ff. (D. Hilbert, Bibliothek des Mathematischen Seminars, Universität Göttingen); from here on "WS 1916/17 Lectures."

$$\sqrt{g}J^{(h)} = -\sum_{\mu,\nu} H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}h^{\mu\nu} + \sum_{\mu,\nu}[\sqrt{g}H]_{\mu\nu}h^{\mu\nu} + D^{(h)}; \tag{42}$$

here

$$[\sqrt{g}H]_{\mu\nu} = \frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_k} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}} \tag{43}$$

is the Lagrangian variational derivative of H, the vanishing of which is the set of gravitational field equations; and:

$$\begin{aligned} D^{(h)} = &\sum_{\mu,\nu,k}\frac{\partial}{\partial w_k}\Big(\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_k}h^{\mu\nu}\Big) + \sum_{\mu,\nu,k,l}\frac{\partial}{\partial w_k}\Big(\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}}h^{\mu\nu}_l\Big) \\ &- \sum_{\mu,\nu,k,l}\frac{\partial}{\partial w_l}\Big(h^{\mu\nu}\frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}}\Big) \end{aligned} \tag{44}$$

i.e. another divergence expression. Obviously, $J^{(h)}$ turns into $J^{(p)}$ if one sets $h^{\mu\nu}$ equal to $p^{\mu\nu}$, thus yielding the desired alternative decomposition:

$$\sqrt{g}J^{(p)} = -\sum_{\mu,\nu} H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} + D^{(h)}\big|_{h=p}. \tag{45}$$

Comparing (45) with (38), it becomes clear that E indeed can be written as a divergence, and thus represents a candidate for the energy expression. In the Proofs this conclusion is presented as one of two properties justifying this designation:

> Call the expression E the energy form. To justify this designation, I prove two properties that the energy form enjoys.
>
> If we substitute the tensor $p^{\mu\nu}$ for $h^{\mu\nu}$ in identity (6) [i.e. (42)] then, taken together with (9) [i.e. (39)] it follows, provided the gravitational equations (8) [i.e. (51) below] are satisfied:
>
> $$E = (D^{(h)})_{h=p} - D^{(p)} \tag{46}$$
>
> or
>
> $$\begin{aligned} E = \sum\Big\{&\frac{\partial}{\partial w_k}\Big(\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_k}g^{\mu\nu}_s p^s\Big) - \frac{\partial}{\partial w_k}\Big(\frac{\partial}{\partial w_l}\Big(\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{kl}}\Big)g^{\mu\nu}_s p^s\Big) \\ &+ \frac{\partial}{\partial w_l}\Big(\sqrt{g}\frac{\partial H}{\partial g^{\mu\nu}_{kl}}g^{\mu\nu}_{sk}p^s\Big)\Big\}, \end{aligned} \tag{47}$$
>
> that is, we have the proposition:
>
> Proposition 1: In virtue of the gravitational equations the energy form E becomes a sum of differential quotients with respect to w_s, that is, it acquires the character of a divergence.[62]

Whereas (47) for an arbitrary H involves an arbitrary combination of electromagnetic and gravitational contributions, Hilbert makes an ansatz $H = K + L$ that allows him to separate these two contributions; in particular, to relate E to his result concerning the energy-momentum tensor of Mie's theory. Accordingly, at this point, he presumably introduces in a missing part of the Proofs (as he does in the corresponding part of Paper 1) the splitting of the Lagrangian (16), and introduces the condition that L not depend on $g_s^{\mu\nu}$.[63] Finally, he writes down explicitly the electromagnetic part of the energy:

> Because K depends only on $g^{\mu\nu}, g_s^{\mu\nu}, g_{lk}^{\mu\nu}$, therefore in ansatz (17) [i.e. (16)], due to (13) [i.e. (47)], the energy E can be expressed solely as a function of the gravitational potentials $g^{\mu\nu}$ and their derivatives, provided L is assumed to depend not on $g_s^{\mu\nu}$, but only on $g^{\mu\nu}, q_s, q_{sk}$. On this assumption, which we shall always make in the following, the definition of the energy (10) [i.e. (39)] yields the expression
>
> $$E = E^{(g)} + E^{(e)}, \tag{48}$$
>
> where the "gravitational energy" $E^{(g)}$ depends only on $g^{\mu\nu}$ and their derivatives, and the "electrodynamic energy" $E^{(e)}$ takes the form
>
> $$E^{(e)} = \sum_{\mu,\nu,s} \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} (g_s^{\mu\nu} p^s - g^{\mu s} p_s^{\nu} - g^{\nu s} p_s^{\mu}), \tag{49}$$
>
> which proves to be *a general invariant multiplied by* $\sqrt{g}$.[64]

(The term in parentheses in equation (49) is $p^{\mu\nu}$, the Lie derivative of the contravariant metric with respect to the vector p^s.)

Hilbert's final expression (49) satisfies what we called his "physical criterion" for finding a definition of the energy since the term $\frac{\partial}{\partial g^{\mu\nu}} \sqrt{g} L$ corresponds—apart from the factor –2 —to the left-hand side of (36), and thus to Mie's energy momentum ten-

---

62 "Der Ausdruck E heiße die Energieform. Um diese Bezeichnung zu rechtfertigen, beweise ich zwei Eigenschaften, die der Energieform zukommen.
Setzen wir in der Identität (6) [i.e. (42)] für $h^{\mu\nu}$ den Tensor $p^{\mu\nu}$ ein, so folgt daraus zusammen mit (9) [(39)], sobald die Gravitationsgleichungen (8) erfüllt sind: [(46); (12) in the original text] or [(47); (13) in the original text] d. h. es gilt der Satz:
Satz 1. Die Energieform E wird vermöge der Gravitationsgleichungen einer Summe von Differentialquotienten nach w_s gleich, d. h. sie erhält Divergenzcharakter." See (Proofs, 6).

63 Compare (Hilbert 1916, 402) with (Proofs, 8), and see the discussion in "Einstein Equations and Hilbert Action …" (in this volume).

64 "Da K nur von $g^{\mu\nu}, g_s^{\mu\nu}, g_{lk}^{\mu\nu}$ abhängt, so läßt sich beim Ansatz (17) die Energie E wegen (13) lediglich als Funktion der Gravitationspotentiale $g^{\mu\nu}$ und deren Ableitungen ausdrücken, sobald wir L nicht von $g_s^{\mu\nu}$, sondern nur von $g^{\mu\nu}, q_s, q_{sk}$ abhängig annehmen. Unter dieser Annahme, die wir im Folgenden stets machen, liefert die Definition der Energie (10) den Ausdruck [(48); (18) in the original text] wo die "Gravitationsenergie" $E^{(g)}$ nur von $g^{\mu\nu}$ und deren Ableitungen abhängt und die "elektrodynamische Energie" $E^{(e)}$ die Gestalt erhält [(49); (19) in the original text] in der sie sich als *eine mit* $\sqrt{g}$ *multiplizierte allgemeine Invariante* erweist." (Proofs, 8)

sor. Hilbert's definition of energy had thus been given a "physical justification" in terms of Mie's theory. But—apart from merely formal similarities—its relation to energy-momentum conservation in classical and special-relativistic theories remains entirely unclear. In the Proofs, as we shall see below, Hilbert's energy expression served still another and even more important function, that of determining admissible coordinate systems.

3.4 Hilbert's Revision of Mie's Program and the Roots of his Leitmotiv in Einstein's Work

Apparently Hilbert was convinced that the relation he established between the variational derivative of the Lagrangian and the energy-momentum tensor (see (36)) singled out Mie's theory as having a special relation to the theory of gravitation.[65] In fact, as we have seen, this conclusion is only justified insofar as one imposes on the electrodynamic term in the Lagrangian the condition that it does not depend on $g_s^{\mu\nu}$. Nevertheless, this result apparently suggested to Hilbert that gravitation may be the more fundamental physical process and that it might be possible to conceive of electromagnetic phenomena as "effects of gravitation."[66] Such an interpretation, which was in line with the reductionist perspective implied by his understanding of the axiomatization of physics, led to a revision of Mie's original aim of basing all of physics on electromagnetism.

In the light of this possibility, the third point of Hilbert's initial research program, the question of the number of independent equations in a generally-covariant theory, must have taken on a new and increased significance. Einstein's hole argument, when applied to Hilbert's formalism, suggests that the fourteen generally-covariant field equations for the 14 gravitational and electromagnetic potentials do not have a unique solution for given boundary values. Consequently, 4 identities must exist between the 14 field equations; and 4 additional, non-covariant equations would be required in order to assure a unique solution; and if these 4 identities were somehow equivalent to the 4 equations for the electromagnetic potentials, then the latter could be considered as a consequence of the 10 gravitational equations by virtue of the unique properties of a generally-covariant variational principle, and Hilbert would indeed be entitled to claim that electromagnetism is an effect of gravitation.

As we have seen, the non-uniqueness of solutions to generally-covariant field equations and the conclusion that such field equations must obey 4 identities, are both issues raised by Einstein in his publications of 1913/14. These writings and his 1915 Göttingen lectures, which Hilbert attended, offered rich sources of information about Einstein's theory. In addition the physicist Paul Hertz, then a participant in the group

65 In fact, this relation between the special-relativistic stress-energy tensor and the variational derivative of the general-relativistic generalization of a Lagrangian giving rise to this stress-energy tensor is quite general, as was pointed out many years later in (Rosenfeld 1940, 1–30; and Belinfante 1939, 887). See also (Vizgin 1989, 304; 1994).

66 See (Proofs, 3) and (Hilbert 1916, 397).

centered around Hilbert in Göttingen, may also have kept Hilbert informed about Einstein's thinking on these issues. For example, in a letter to Hertz of August 1915, Einstein raised the problem of solving hyperbolic partial differential equations for arbitrary boundary values and discussed the necessity of introducing four additional equations to restore causality for a set of generally-covariant field equations.[67]

Einstein's treatment of these issues thus forms the background to the crucial theorem, on which Hilbert's entire approach is based, his *Leitmotiv*, labelled "Theorem I" in the Proofs:

> The guiding motive for setting up the theory is given by the following theorem, the proof of which I shall present elsewhere.
>
> Theorem I. If J is an invariant under arbitrary transformations of the four world parameters, containing n quantities and their derivatives, and if one forms from
>
> $$\delta\int J\sqrt{g}d\tau = 0 \tag{50}$$
>
> the n variational equations of Lagrange with respect to each of the n quantities, then in this invariant system of n differential equations for the n quantities there are always four that are a consequence of the remaining $n-4$ —in the sense that, among the n differential equations and their total derivatives, there are always four linear and mutually independent combinations that are satisfied identically.[68]

For a Lagrangian H depending on the gravitational and the electrodynamic potentials and their derivatives, Hilbert derived 10 field equations for the gravitational potentials $g^{\mu\nu}$ and 4 for the electrodynamic potentials q_s from such a variational principle (50):

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_k} - \sum_{k,l}\frac{\partial^2}{\partial w_k\partial w_l}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}}, \quad (\mu,\nu = 1,2,3,4), \tag{51}$$

$$\frac{\partial\sqrt{g}H}{\partial q_h} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}}, \quad (h = 1,2,3,4). \tag{52}$$

---

67 Einstein to Paul Hertz, 22 August 1915, (CPAE 8, 163–164). See (Howard and Norton 1993) for an extensive historical discussion.

68 "Das Leitmotiv für den Aufbau der Theorie liefert der folgende mathematische Satz, dessen Beweis ich an einer anderen Stelle darlegen werde.
Theorem I. Ist J eine Invariante bei beliebiger Transformation der vier Weltparameter, welche n Größen und ihre Ableitungen enthält, und man bildet dann aus [(50)] in Bezug auf jene n Größen die n Lagrangeschen Variationsgleichungen, so sind in diesem invarianten System von n Differentialgleichungen für die n Größen stets vier eine Folge der $n-4$ übrigen — in dem Sinne, daß zwischen den n Differentialgleichungen und ihren totalen Ableitungen stets vier lineare, von einander unabhängige Kombinationen identisch erfüllt sind." (Proofs, 2–3) See (Hilbert 1916, 396–397). See (Rowe 1999) for a discussion of the debate on Hilbert's Theorem I among Göttingen mathematicians.

In both the Proofs and Paper 1, Hilbert erroneously claimed that one can consider the last four equations to be a consequence of the 4 identities that must hold, according to his Theorem I, between the 14 differential equations:

> Let us call equations (4) [i.e. (51)] the fundamental equations of gravitation, and equations (5) [i.e. (52)] the fundamental electrodynamic equations, or generalized Maxwell equations. Due to the theorem stated above, the four equations (5) [i.e. (52)] can be viewed as a consequence of equations (4) [i.e. (51)]; that is, because of that mathematical theorem we can immediately assert the claim *that in the sense explained above electrodynamic phenomena are effects of gravitation*. I regard this insight as the simple and very surprising solution of the problem of Riemann, who was the first to search for a theoretical connection between gravitation and light.[69]

We shall come back to this claim later, in connection with Hilbert's proof of a special case of Theorem I.

The fact that Hilbert did not give a proof of this theorem makes it difficult to assess its heuristic roots. No doubt, of course, some of these roots lay in Hilbert's extensive mathematical knowledge, in particular, of the theory of invariants. But the lack of a proof in Paper 1, as well as the peculiar interpretation of it in the Proofs, make it plausible that the theorem also had roots in Einstein's hole argument on the ambiguity of solutions to generally-covariant field equations.

In fact, in the Proofs, Hilbert placed the implications of Theorem I for his field theory in the context of the problem of causality, as Einstein had done for the hole argument. But while the hole argument was formulated in terms of a boundary value problem for a closed hypersurface, Hilbert posed the question of causality in terms of an initial value problem for an open one, thus adapting it to Cauchy's theory of systems of partial differential equations:

> Since our mathematical theorem shows that the axioms I and II [essentially amounting to the variational principle (50), see the discussion below] considered so far can produce only ten essentially independent equations; and since, on the other hand, if general invariance is maintained, more than ten essentially independent equations for the 14 potentials $g_{\mu\nu}, q_s$ are not at all possible; therefore—provided that we want to retain the determinate character of the basic equation of physics corresponding to Cauchy's theory of differential equations— the demand for four further non-invariant equations in addition to (4) [i.e. (51)] and (5) [i.e. (52)] is imperative.[70]

Hilbert's counting of needed equations closely parallels Einstein's: the number of field equations (10 in Einstein's case and 14 in Hilbert's) plus 4 coordinate restrictions to make sure that causality is preserved. Since Hilbert, in contrast to Einstein,

---

69 "Die Gleichungen (4) mögen die Grundgleichungen der Gravitation, die Gleichungen (5) die elektrodynamischen Grundgleichungen oder die verallgemeinerten Maxwellschen Gleichungen heißen. Infolge des oben aufgestellten Theorems können die vier Gleichungen (5) als eine Folge der Gleichungen (4) angesehen werden, d. h. wir können unmittelbar wegen jenes mathematischen Satzes die Behauptung aussprechen, *daß in dem bezeichneten Sinne die elektrodynamischen Erscheinungen Wirkungen der Gravitation sind*. In dieser Erkenntnis erblicke ich die einfache und sehr überraschende Lösung des Problems von Riemann, der als der Erste theoretisch nach dem Zusammenhang zwischen Gravitation und Licht gesucht hat." (Proofs, 3; Hilbert 1916, 397–398)

had started from a generally-covariant variational principle, he obtained, in addition, 4 identities that, he claimed, imply the electrodynamic equations (52).

Additional evidence for our conjecture that Einstein's hole argument was one of the roots of Hilbert's theorem (and thus of its later elaboration by Emmy Noether) is provided by other contemporary writings of Hilbert, which will be discussed below in connection with Hilbert's second paper, in which the problem of causality is addressed explicitly.[71]

3.5 Energy-Momentum Conservation and Coordinate Restrictions

As we shall see in this section, the Proofs show that Hilbert was convinced that causality requires four supplementary non-covariant equations to fix the admissible coordinate systems. In identifying these coordinate restrictions, he again followed closely in Einstein's tracks. As did the latter, Hilbert invoked energy-momentum conservation in order to justify physically the choice of a preferred reference frame. After formulating his version of energy-momentum conservation, he introduced the following axiom:

> Axiom III (axiom of space and time). *The spacetime coordinates are those special world parameters for which the energy theorem* (15) [i.e. (57) below] *is valid*.
>
> According to this axiom, space and time in reality provide a special labeling of the world's points such that the energy theorem holds.
>
> Axiom III implies the existence of equations (16) [$d^{(g)}\sqrt{g}H / dw_s = 0$]: these four differential equations (16) complete the gravitational equations (4) [i.e. (51)] to give a system of 14 equations for the 14 potentials $g_{\mu\nu}, q_s$, *the system of fundamental equations of physics*. Because of the agreement in number between equations and potentials to be determined, the principle of causality for physical processes is also guaranteed, revealing to us the closest connection between the energy theorem and the principle of causality, since each presupposes the other.[72]

The strategy Hilbert followed to extract these coordinate restrictions from the requirement of energy conservation closely followed that of Einstein's *Entwurf* theory of 1913/14. Even before he developed the hole argument, energy-momentum conservation played a crucial role in justifying the lack of general covariance of his

---

70 "Indem unser mathematisches Theorem lehrt, daß die bisherigen Axiome I und II für die 14 Potentiale nur zehn wesentlich von einander unabhängige Gleichungen liefern können, andererseits bei Aufrechterhaltung der allgemeinen Invarianz mehr als zehn wesentlich unabhängige Gleichungen für die 14 Potentiale $g_{\mu\nu}, q_s$ garnicht möglich sind, so ist, wofern wir der Cauchyschen Theorie der Differentialgleichungen entsprechend den Grundgleichungen der Physik den Charakter der Bestimmtheit bewahren wollen, die Forderung von vier weiteren zu (4) und (5) hinzutretenden nicht invarianten Gleichungen unerläßlich." (Proofs, 3–4)

71 See, e.g., his SS 1916 Lectures, in particular p. 108, as well as an undated typescript preserved at Göttingen, in SUB Cod. Ms. 642, entitled *Das Kausalitätsprinzip in der Physik*, henceforth cited as the "Causality Lecture." Page 4 of this typescript, describing a construction equivalent to Einstein's hole argument, is discussed below.

gravitational field equations. He was convinced that energy-momentum conservation actually required a restriction of the covariance group.[73] An the beginning of 1914, after having formulated the hole argument, he described the connection between coordinate restrictions and energy-momentum conservation in the *Entwurf* theory as follows:

> Once we have realized that an acceptable theory of gravitation necessarily implies a specialization of the coordinate system, it is also easily seen that the gravitational equations given by us are based upon a special coordinate system. Differentiation of equations (II) with respect to x_ν [the field equations in the form
>
> $$\sum_{\alpha\beta\mu} \frac{\partial}{\partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta}g_{\sigma\mu}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) = \kappa(\mathfrak{T}_{\sigma\nu} + \mathfrak{t}_{\sigma\nu})]$$
>
> and summation over ν, and taking into account equations (III), [the conservations equations in the form
>
> $$\sum_{\nu} \frac{\partial}{\partial x_\nu}(\mathfrak{T}_{\sigma\nu} + \mathfrak{t}_{\sigma\nu}) = 0\ ] \tag{53}$$
>
> yields the relations (IV)
>
> $$\left[\sum_{\alpha\beta\mu\nu} \frac{\partial^2}{\partial x_\nu \partial x_\alpha}\left(\sqrt{-g}\gamma_{\alpha\beta}g_{\sigma\mu}\frac{\partial \gamma_{\mu\nu}}{\partial x_\beta}\right) = 0\right], \tag{54}$$
>
> that is, four differential conditions for the quantities $g_{\mu\nu}$, which we write in the abbreviated form
>
> $$B_\sigma = 0. \tag{55}$$
>
> These quantities B_σ do not form a generally-covariant vector, as will be shown in §5. From this one can conclude that the equations $B_\sigma = 0$ represent a real restriction on the choice of coordinate system.[74]

In a later 1914 paper, Einstein discussed the physical significance and the transformation properties of the gravitational energy-momentum term $\mathfrak{t}_\sigma^\nu$:

> According to the considerations of §10, the equations (42 c) [i.e. (53)] represent the conservation laws of momentum and energy for matter and gravitational field combined. The $\mathfrak{t}_\sigma^\nu$ are those quantities, related to the gravitational field, which are analogies in physical

---

72 "Axiom III (Axiom von Raum und Zeit). *Die Raum-Zeitkoordinaten sind solche besonderen Weltparameter, für die der Energiesatz (15) gültig ist*.
Nach diesem Axiom liefern in Wirklichkeit Raum und Zeit eine solche besondere Benennung der Weltpunkte, daß der Energiesatz gültig ist.
Das Axiom III hat das Bestehen der Gleichungen (16) zur Folge: diese vier Differentialgleichungen (16) vervollständigen die Gravitationsgleichungen (4) zu einem System von 14 Gleichungen für die 14 Potentiale $g^{\mu\nu}$, q_s: *dem System der Grundgleichungen der Physik*. Wegen der Gleichzahl der Gleichungen und der zu bestimmenden Potentiale ist für das physikalische Geschehen auch das Kausalitätsprinzip gewährleistet, und es enthüllt sich uns damit der engste Zusammenhang zwischen dem Energiesatz und dem Kausalitätsprinzip, indem beide sich einander bedingen." (Proofs, 7)

73 See, e.g., (Einstein 1913, 1258).

> interpretation to the components $\mathfrak{T}_\sigma^\nu$ of the energy tensor (V-Tensor) [i.e. tensor density]. It is to be emphasized that the $\mathfrak{t}_\sigma^\nu$ do not have tensorial covariance under arbitrary admissible [coordinate] transformations but only under linear transformations. Nevertheless, we call ($\mathfrak{t}_\sigma^\nu$) the energy tensor of the gravitational field.[75]

Similarly, Hilbert notes that his energy-form is invariant with respect to linear transformations; he shows that E can be decomposed with respect to the vector p^j as follows (Proofs, 6):

$$E = \sum_s e_s p^s + \sum_{s,l} e_s^l p_l^s \tag{56}$$

where e_s and e_s^l are independent of p^j. If one compares this expression with Einstein's (53), then the analogy between the two suggests that the two-index object e_s^l should play the same role in Hilbert's theory as does the total energy-momentum tensor in Einstein's theory, satisfying a divergence equation of the form:

$$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0. \tag{57}$$

Hilbert shows that this equation holds only if e_s vanishes, in which case:

$$E = \sum_{s,l} e_s^l p_l^s. \tag{58}$$

This equation can be related to energy conservation; Hilbert calls this the "normal form" of the energy. The fact that the last two equations imply each other was, for Hilbert, apparently a decisive reason for calling E the energy form. Indeed, this equivalence is the subject of his second theorem about the energy-form. Although the relevant part of the Proofs is missing,[76] Hilbert's theorem and its proof can be reconstructed:

---

74 "Nachdem wir so eingesehen haben, daß eine brauchbare Gravitationstheorie notwendig einer Spezialisierung des Koordinatensystems bedarf, erkennen wir auch leicht, daß bei den von uns angegebenen Gravitationsgleichungen ein spezielles Koordinatensystem zugrunde liegt. Aus den Gleichungen (II) folgen nämlich durch Differentiation nach x_ν und Summation über ν unter Berücksichtigung der Gleichungen (III) die Beziehungen (IV) also vier Differentialbedingungen für die Größen $g_{\mu\nu}$, welche wir abgekürzt $B_\sigma = 0$ schreiben wollen.
Diese Größen B_σ bilden, wie in §5 gezeigt ist, keinen allgemein-kovarianten Vektor. Hieraus kann geschlossen werden, daß die Gleichungen $B_\sigma = 0$ eine wirkliche Bedingung für die Wahl des Koordinatensystems darstellen." (Einstein and Grossmann 1914, 218–219)

75 "Die Gleichungen (42 c) drücken nach den in §10 gegebenen Überlegungen die Erhaltungssätze des Impulses und der Energie für Materie und Gravitationsfeld zusammen aus. $\mathfrak{t}_\sigma^\nu$ sind diejenigen auf das Gravitationsfeld bezüglichen Größen, welche den Komponenten $\mathfrak{T}_\sigma^\nu$ des Energietensors (V-Tensors) [i.e. tensor density] der physikalischen Bedeutung nach analog sind. Es sei hervorgehoben, daß die $\mathfrak{t}_\sigma^\nu$ nicht beliebigen berechtigten, sondern nur linearen Transformationen gegenüber Tensorkovarianz besitzen; trotzdem nennen wir ($\mathfrak{t}_\sigma^\nu$) den Energietensor des Gravitationsfeldes." (Einstein 1914b, 1077)

76 The top portion of the Proofs, p. 7, is missing.

Theorem 2 must have asserted that:

$$e_s = \frac{\partial e_s^{\ l}}{\partial w_l} \tag{59}$$

This assertion is easily proven by following the lines indicated in the surviving portion of Hilbert's argument. From (38) and (56) it follows that:

$$\sqrt{g}J^{(p)} + H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} = e_s p^s + e_s^l p_l^s + D^{(p)}, \tag{60}$$

which can be rewritten as:

$$\sqrt{g}J^{(p)} + H\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}}p^{\mu\nu} = \left(e_s - \frac{\partial e^{l}_{\ s}}{\partial w_l}\right)p^s + \overline{D^{(p)}}, \tag{61}$$

where $\overline{D^{(p)}}$ is still a divergence. If now the integral over a region Ω, on the boundary of which p^s and its first derivative vanish, is taken on both sides, then the surface terms vanish. Thus one obtains in view of (42):

$$\int_\Omega [\sqrt{g}H]_{\mu\nu}p^{\mu\nu}dx^4 = \int_\Omega \left(e_s - \frac{\partial e^{l}_{\ s}}{\partial w_l}\right)p^s(dx^4). \tag{62}$$

But the left-hand side vanishes when the gravitational field equations hold, and p^s is an arbitrary vector field, from which (59) follows.

Theorem 2 provides Hilbert with the desired coordinate restrictions:

> This theorem shows that the divergence equation corresponding to the energy theorem of the old theory
>
> $$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0 \tag{63}$$
>
> holds if and only if the four quantities e_s vanish ... .[77]

After these preparations, Hilbert introduces Axiom III, quoted at the beginning of this section, which establishes a distinction between the arbitrary world parameters w_l and the restricted class of coordinates that constitute "a spacetime reference system." In fact, the latter are those world parameters satisfying the coordinate restrictions $e_s = 0$ following from Hilbert's energy condition. In analogy to the "justified coordinate transformations" of Einstein's 1913/14 theory leading from one "adapted

77 "Dieser Satz zeigt, daß die dem Energiesatz der alten Theorie entsprechende Divergenzgleichung [(63); (15) in the original text] dann und nur dann gelten kann, wenn die vier Größen e_s verschwinden ..." (Proofs, 7).

coordinate system" to another, Hilbert introduced spacetime transformations that lead from one "normal form" of the energy to another:

> To the transition from one spacetime reference system to another one corresponds the transformation of the energy form from one so-called "normal form"
>
> $$E = \sum_{s,l} e_s^l p_l^s \tag{64}$$
>
> to another normal form.[78]

The claim that Hilbert's introduction of coordinate restrictions was guided by the goal of recovering the ordinary divergence form of energy-momentum conservation is supported by his later use of this argument in a discussion with Felix Klein. In a letter to Hilbert, Klein recounted how, at a meeting of the Göttingen Academy, he had argued that, for the energy balance of a field, one should take into account only the energy tensor of matter (including that of the electromagnetic field) without ascribing a separate energy-momentum tensor to the gravitational field.[79] This suggestion was taken up by Carl Runge, who had given an expression for energy-momentum conservation that, in his letter to Hilbert, Klein called "regular" and found similar to what happens in the "elementary theory."[80] Starting from an expression for the covariant divergence of the stress-energy tensor:

$$\sum_{\mu\nu}\left(\sqrt{g}T_{\mu\nu}g_{\sigma}^{\mu\nu} + 2\frac{\partial}{\partial w_\nu}(\sqrt{g}T_{\mu\sigma}g^{\mu\nu})\right) = 0 \qquad \sigma = 1, 2, 3, 4 \tag{65}$$

Runge obtained his "regular" expression by imposing the four equations:

$$\sum_{\mu\nu}\sqrt{g}T_{\mu\nu}g_{\sigma}^{\mu\nu} = 0, \tag{66}$$

thus specifying a preferred class of coordinate systems. In his response, Hilbert sent Klein three pages of the Proofs to show that he had anticipated Runge's line of reasoning:

> I send you herewith my first proofs [footnote: Please kindly return these to me as I have no other record of them.] (3 pages) of my first communication, in which I also implemented Runge's ideas; in particular with theorem 1, p. 6, in which the divergence character of the energy is proven. I later omitted the whole thing as the thing did not seem to me to be fully mature. I would be very pleased if progress could now be made. For this it is necessary to retrieve the old energy conservation laws in the limiting case of Newtonian theory.[81]

---

78 "Dem Übergang von einem Raum-Zeit-Bezugssystem zu einem anderen entspricht die Transformation der Energieform von einer sogenannten "Normalform" [(64)] auf eine andere Normalform." (Proofs, 7)

79 Felix Klein to David Hilbert, 5 March 1918, (Frei 1985, 142–143).

80 For a discussion of Runge's work, see (Rowe 1999).

Hilbert's final sentence confirms that the recovery of the familiar form of energy conservation was his goal. However, at the time of the Proofs, it was clearly not his aim to eliminate the energy-momentum expression of the gravitational field from the energy balance, as the above reference to Runge might suggest. On the contrary, as we have seen above (see (48)), Hilbert followed Einstein in attempting to treat the contributions to the total energy from the electromagnetic and the gravitational parts on an equal footing.

In summary, Hilbert's first steps in the realization of his research program were the derivation of what he regarded as the unique relation between the variational derivative of Mie's Lagrangian and Mie's energy momentum tensor, and the formulation of a theorem, by means of which he hoped to show that the electromagnetic field equations follow from the gravitational ones. Albeit problematic from a modern perspective, these steps become understandable in the context of Hilbert's application of his axiomatic approach to Einstein's non-covariant theory of gravitation and Mie's theory of matter. These first steps in turn shaped Hilbert's further research. They effected a change of perspective from viewing electrodynamics and gravitation on an equal footing to his vision of deriving electromagnetism from gravitation. As a consequence, the structure of Hilbert's original, non-covariant theory, in spite of the covariance of Hilbert's gravitational equations and the different physical interpretation that he gave to his equations, is strikingly similar to that of Einstein's 1913/14 *Entwurf* theory of gravitation.

3.6 Electromagnetism as an Effect of Gravitation: The Core of Hilbert's Theory

Now we come to the part of Hilbert's program that today is often considered to contain his most important contributions to general relativity: the contracted Bianchi identities and a special case of Noether's theorem. We shall show that, in the original version of Hilbert's theory, these mathematical results actually constituted part of a different physical framework that also affected their interpretation. In a later section, we shall see how these results were transformed, primarily due to the work of Hendrik Antoon Lorentz and Felix Klein, into constituents of general relativity. In the hindsight of general relativity, it appears as if Hilbert first derived the contracted Bianchi identities, applied them to the gravitational field equations with an electromagnetic source-term, and then showed that the electrodynamic variables necessarily satisfy the Maxwell equations. This last result, however, is valid only under addi-

81 "Anbei schicke ich Ihnen meine erste Korrektur [footnote: Bitte dieselbe mir wieder freundlichst zustellen zu wollen, da ich sonst keine Aufzeichnungen habe.] (3 Blätter) meiner ersten Mitteilung, in der ich gerade die Ideen von Runge auch ausgeführt hatte; insbesondere auch mit Satz l, S. 6, in dem der Divergenzcharakter der Energie bewiesen wird. Ich habe aber die ganze Sache später unterdrückt, weil die Sache mir nicht reif erschien. Ich würde mich sehr freuen, wenn jetzt der Fortschritt gelänge. Dazu ist aber nötig im Grenzfalle zur Newtonschen Theorie die alten Energiesätze wiederzufinden." Tilman Sauer suggested that the pages sent to Klein were the three sheets of the Proofs bearing Roman numbers I, II, and III, see (Sauer 1999, 544).

tional assumptions that run counter to Mie's program. From the point of view of general relativity, Hilbert obtained Maxwell's equations as a consequence of the integrability conditions for the gravitational field equations with electromagnetic source term, as if he had treated a special case of Einstein's equations and expressed certain of their general properties in terms of this special case. From Hilbert's point of view, however, he had derived the electrodynamic equations as a consequence of the gravitational ones; his derivation was closely interwoven with other results of his theory that pointed to electromagnetism as an effect of gravitation. For him, the equation, on the basis of which he argued that electrodynamics is a consequence of gravitation, was a result of four ingredients, two of which are other links between gravitation and electrodynamics, and all of which are based on his generally-covariant variational principle:

- a general theorem corresponding to the contracted Bianchi identities,
- the field equations following from the variational principle,
- the relation between Mie's energy-momentum tensor and the variational derivative of the Lagrangian, and
- the way in which the derivatives of the electrodynamic potentials enter Mie's Lagrangian.

In the Proofs, the general theorem is:

> Theorem III. If J is an invariant depending only on the $g^{\mu\nu}$ and their derivatives and if, as above, the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$ are denoted by $[\sqrt{g}J]_{\mu\nu}$, then the expression — in which $h^{\mu\nu}$ is understood to be any contravariant tensor —
>
> $$\frac{1}{\sqrt{g}}\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}h^{\mu\nu} \tag{67}$$
>
> represents an invariant; if in this sum we substitute in place of $h^{\mu\nu}$ the particular tensor $p^{\mu\nu}$ and write
>
> $$\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}p^{\mu\nu} = \sum_{s,l}(i_s p^s + i_s^l p_l^s), \tag{68}$$
>
> where then the expressions
>
> $$\begin{aligned} i_s &= \sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}g_s^{\mu\nu}, \\ i_s^l &= -2\sum_{\mu}[\sqrt{g}J]_{\mu s}g^{\mu l} \end{aligned} \tag{69}$$
>
> depend only on the $g^{\mu\nu}$ and their derivatives, then we have
>
> $$i_s = \sum_l \frac{\partial i_s^l}{\partial w_l} \tag{70}$$
>
> in the sense, that this equation is identically fulfilled for all arguments, that is for the $g^{\mu\nu}$ and their derivatives.[82]

Here, (68) follows from an explicit calculation taking into account the definition of $p^{\mu\nu}$; the identity (70) follows if in analogy to (61) one rewrites (68) as:

$$[\sqrt{g}J]_{\mu\nu}p^{\mu\nu} = \left(i_s - \frac{\partial i_s^l}{\partial w_l}\right)p^s + \frac{\partial}{\partial w_l}(i_s^l p^s), \tag{71}$$

and, as in the earlier derivation, carries out a surface integration. Theorem III, in the form of (70), thus corresponds to the contracted Bianchi identities.

Hilbert next applies Theorem III to the Lagrangian $H = K + L$ using his knowledge about its electrodynamic part (see the last two "ingredients" listed above) in order to extract the electrodynamic equations from the identity for L that corresponds to (70). From a modern point of view, it is remarkable that Hilbert did not consider the physical significance of this identity for the gravitational part K of the Lagrangian, but only for the electrodynamic part. For Hilbert, however, this was natural; presumably he was convinced, on the basis of Theorem I, that generally-covariant equations for gravitation are impossible as a "stand-alone" theory. Consequently, it simply made no sense to interpret the gravitational part of these equations by itself.

Assuming the split of the Lagrangian into $K + L$, the gravitational and electrodynamic parts as in (16), he rewrites (51) as:[83]

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{72}$$

He next applies (69) to the invariant K:

$$i_s = \sum_{\mu,\nu}[\sqrt{g}K]_{\mu\nu}g_s^{\mu\nu}, \tag{73}$$

and

$$i_s^l = -2\sum_{\mu}[\sqrt{g}K]_{\mu s}g^{\mu l}, \quad (\mu = 1, 2, 3, 4). \tag{74}$$

From the modern point of view, it would be natural to invoke the identity (70) in order to derive its implications for the source term of the gravitational field equations, i.e., the second term of (72) in Hilbert's notation. In this way, one would obtain an integrability condition for the gravitational field equations that can be interpreted as representing energy-momentum conservation.

---

82 "Theorem III. Wenn J eine nur von den $g^{\mu\nu}$ und deren Ableitungen abhängige Invariante ist, und, wie oben, die Variationsableitungen von $\sqrt{g}J$ bezüglich $g^{\mu\nu}$ mit $[\sqrt{g}J]_{\mu\nu}$ bezeichnet werden, so stellt der Ausdruck — unter $h^{\mu\nu}$ irgend einen kontravarianten Tensor verstanden — [(67)] eine Invariante dar; setzen wir in dieser Summe an Stelle von $h^{\mu\nu}$ den besonderen Tensor $p^{\mu\nu}$ ein und schreiben [(68)] wo alsdann die Ausdrücke [(69)] lediglich von den $g^{\mu\nu}$ und deren Ableitungen abhängen, so ist [(70)] in der Weise, daß diese Gleichung identisch für alle Argumente, nämlich die $g^{\mu\nu}$ und deren Ableitungen, erfüllt ist." (Proofs, 9; Hilbert 1916, 399)

83 See (Proofs, 11; Hilbert 1916, 405).

Hilbert proceeded differently, using Theorem III to further elaborate what he considered his crucial insight into the relation between Mie's energy-momentum tensor and the variational derivative of L. Consequently he focussed on (36), from which he attempted to extract the equations for the electromagnetic field. In fact, the left-hand side of this equation can (in view of (72) and (74)) be rewritten as $-i_\nu^m$. Consequently, differentiating the right-hand side of (36) with respect to w_m and summing over m, Theorem III yields:

$$\begin{aligned} i_\nu &= \sum_m \frac{\partial}{\partial w_m}\Big(-\sqrt{g}L\delta_\nu^m + \frac{\partial\sqrt{g}L}{\partial q_m}q_\nu + \sum_s \frac{\partial\sqrt{g}L}{\partial M_{sm}}M_{s\nu}\Big) \\ &= -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \Bigg\{ q_\nu \frac{\partial}{\partial w_m}\Big([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\Big) \\ &\quad + q_{\nu m}\Big([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\Big)\Bigg\} \\ &\quad + \sum_s \Big([\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}\Big)M_{s\nu} + \sum_{s,m}\frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}, \end{aligned} \tag{75}$$

where use has been made of:

$$\frac{\partial\sqrt{g}L}{\partial q_m} = [\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} \tag{76}$$

and

$$-\sum_m \frac{\partial}{\partial w_m}\frac{\partial\sqrt{g}L}{\partial q_{sm}} = [\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}. \tag{77}$$

Here $[\sqrt{g}L]_h$ denotes the Lagrangian derivative of $\sqrt{g}L$ with respect to the electrodynamic potentials q_h:

$$[\sqrt{g}L]_h = \frac{\partial\sqrt{g}L}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}L}{\partial q_{hk}}; \tag{78}$$

the vanishing of which constitutes the electromagnetic field equations. At this point Hilbert makes use of the last ingredient, the special way in which the derivatives of the potentials enter Mie's Lagrangian. Taking into account (27), one obtains:

$$\sum_{m,s}\frac{\partial^2}{\partial w_m\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} = 0, \tag{79}$$

so that (75) can be rewritten as:

$$i_\nu = -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \left(q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m + M_{m\nu}[\sqrt{g}L]_m \right) + \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} + \sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}} \frac{\partial M_{s\nu}}{\partial w_m}. \tag{80}$$

While the right-hand side of this equation only involves the electrodynamic part of the Lagrangian, in view of (73) this is not the case for the left-hand side. Therefore, Hilbert once more uses the field equations, in the form of (72), for i_ν to obtain an expression entirely in terms of the electrodynamic part of the Lagrangian. For this purpose, he first writes:

$$-\frac{\partial\sqrt{g}L}{\partial w_\nu} = -\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial g^{sm}} g^{sm}_\nu - \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} - \sum_{m,s} \frac{\partial\sqrt{g}L}{\partial q_{ms}} \frac{\partial q_{ms}}{\partial w_\nu}, \tag{81}$$

and then uses (72) and (73) to identify the first term on the right-hand side as i_ν. Hilbert thus reaches his goal of transforming the identity following from Theorem III into an equation involving only the electromagnetic potentials. A further simplification results from noting that the last term on the right-hand side of (81) is, apart from its sign, identical to the last term of (80). (This is because:

$$\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}} \left(\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu} \right) = 0, \tag{82}$$

which follows from the definition (28) of M_{sm}.)

Finally, using (80), Hilbert obtains:

$$\sum_m \left(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m \right) = 0. \tag{83}$$

Summarizing what he had achieved, Hilbert claimed:

> ... from the gravitational equations (4) [i.e. (51)] there follow indeed the four linearly independent combinations (32) [i.e. (83)] of the basic electrodynamic equations (5) [i.e. (52)] and their first derivatives. *This is the entire mathematical expression of the general claim made above about the character of electrodynamics as an epiphenomenon of gravitation*.[84]

On closer inspection, Hilbert's claim turns out to be problematic. One might try to interpret it in either of two ways: the electromagnetic field equations follow either differentially or algebraically from (83).

---

84 "... aus den Gravitationsgleichungen (4) folgen in der Tat die vier von einander unabhängigen linearen Kombinationen (32) der elektrodynamischen Grundgleichungen (5) und ihrer ersten Ableitungen. *Dies ist der ganze mathematische Ausdruck der oben allgemein ausgesprochenen Behauptung über den Charakter der Elektrodynamik als einer Folgeerscheinung der Gravitation*." (Proofs, 12) In (Hilbert 1916, 406), "ganze" [entire] is corrected to "genaue" [exact] in the last sentence.

In the first case one would have to show that, if these equations hold on an initial hypersurface $w_4 = const$, then they hold everywhere off that hypersurface by virtue of the identities (83). Indeed it follows from these identities that, if these equations hold on $w_4 = 0$:

$$\frac{\partial[\sqrt{g}L]_4}{\partial w_4} = 0, \tag{84}$$

so that, by iteration, $[\sqrt{g}L]_4 = 0$ holds everywhere provided that it holds initially and that the other three field equations hold everywhere. But the time derivatives of the other three field equations,

$$\frac{\partial[\sqrt{g}L]_m}{\partial w_4} \qquad m = 1, 2, 3 \tag{85}$$

remain unrestricted by the identity so that one cannot simply give the electromagnetic field equations on an initial hypersurface and have them continue to hold automatically off it as a consequence of (83).

In the second case, it is clear that the field equations can only hold algebraically by virtue of (83) if the second term vanishes; this implies that the theory is gauge invariant, i.e. that the potentials themselves do not enter the field equations. In that case one indeed obtains an additional identity from gauge invariance:

$$\frac{\partial[\sqrt{g}L]_m}{\partial w_m} = 0. \tag{86}$$

(In the usual Maxwell theory this is the identity that guarantees conservation of the charge-current vector.) However, this cannot have been the argument Hilbert had in mind when stating his claim. First of all, he did not introduce the additional assumptions required—and could not have introduced them because they violated his physical assumptions;[85] and second he did not derive the identity for gauge-invariant electromagnetic Lagrangians that makes this argument work. As illustrated by Klein's later work, the derivation of these identities is closely related to a different perspective on Hilbert's results, a perspective in which electromagnetism is no longer, as in Hilbert's Proofs, treated as an epiphenomenon of gravitation, but in which both are treated in parallel.[86]

In summary, Hilbert's claim that the electromagnetic equations are a consequence of the gravitational ones turns out to be an interpretation forced upon his mathematical results by his overall program rather than being implied by them. In any case, this

85 Mie's original theory is in fact not gauge invariant, and in the version adopted by Hilbert one of the invariants involves a function of the electromagnetic potential vector, see (33).

86 Compare Klein's attempt to derive analogous equations for the gravitational and the electromagnetic potentials, from which the Maxwell equations then are derived, (Klein 1917, 472–473).

interpretation is different from that given to the corresponding results in general relativity and usually associated with Hilbert's work.

3.7 The Deductive Structure of the Proofs Version

Having attempted to reconstruct the line of reasoning Hilbert followed while developing the original version of his theory, we now summarize the way in which he presented these results in the Proofs. This serves as a review of the deductive structure of his theory, indicating which results were emphasized by Hilbert, and facilitating a comparison between the Proofs and the published versions.

We begin by recalling the elements of this deductive structure that Hilbert introduced explicitly:

- Axiom I "Mie's Axiom von der Weltfunktion," (see (19))
- Axiom II "Axiom von der allgemeinen Invarianz," (see the passage below (19))
- Axiom III "Axiom von Raum und Zeit," (see the passage above (55))
- Theorem I, Hilbert's *Leitmotiv*, (see (50))
- Theorem II, Lie derivative of the Lagrangian, (see (23))
- Theorem III, contracted Bianchi identities, (see (70))
- Proposition 1, divergence character of the energy expression, (see (47))
- Proposition 2, identity obeyed by the components of the energy expression, (see (59)).

He also used the following assumptions, introduced as part of his deductive structure without being explicitly stated:

- vanishing of the divergence of the energy expression (see (63))
- splitting of the Lagrangian into gravitational and electrodynamical terms (see (16))
- the assumption that the electrodynamical term does not depend on the derivatives of the metric tensor (see (25)).

There are, furthermore, the following physical results, not labelled as theorems:

- the field equations (see (51) and (52))
- the energy expression (see (39)) and the related coordinate restrictions (see (63))
- the form of Mie's Lagrangian (see (27))
- the relation between Mie's energy tensor and Lagrangian (see (36))
- the relation between the electromagnetic and gravitational field equations (see (83)).

The exposition of Hilbert's theory in the Proofs can be subdivided into four sections, to which we give short titles and list under each the relevant elements of his theory:

1. Basic Framework (Proofs, 1–3)

   Axioms I and II, Theorem I, and the field equations for gravitation and electromagnetism

2. Causality and the Energy Expression (Proofs, 3–8)
   the energy expression, Propositions 1 and 2, the divergence character of the energy expression, Axiom III, the coordinate restrictions, the split of the Lagrangian into gravitational and electrodynamical terms, and the structure of the electrodynamical term
3. Basic Theorems (Proofs, 8–9)
   Theorems II and III
4. Implications for Electromagnetism (Proofs, 9–13)
   the form of Mie's Lagrangian, its relation to his energy tensor, and the relation between electromagnetic and gravitational field equations.

The sequence in which Hilbert presented these elements suggests that he considered its implications for electromagnetism as the central results of his theory. Indeed, the gravitational field equations are never explicitly given and only briefly considered at the beginning as part of the general framework, whereas the presentation concludes with three results concerning Mie's theory. The centrality of these electromagnetic implications for him is also clear from his introductory and concluding remarks. Hilbert's initial discussion mentions Mie's electrodynamics first, and closes with the promise of further elaboration of the consequences of his theory for electrodynamics:

> The far reaching ideas and the formation of novel concepts by means of which Mie constructs his electrodynamics, and the prodigious problems raised by Einstein, as well as his ingeniously conceived methods of solution, have opened new paths for the investigation into the foundations of physics.
>
> In the following—in the sense of the axiomatic method—I would like to develop from three simple axioms a new system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.[87]

In his conclusion, Hilbert makes clear what he had in mind here: a solution of the riddles of atomic physics:

> As one can see, the few simple assumptions expressed in axioms I, II, III suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense called for by Einstein, but I am also convinced that through the basic equations established here the most

87 "Die tiefgreifenden Gedanken und originellen Begriffsbildungen vermöge derer Mie seine Elektrodynamik aufbaut, und die gewaltigen Problemstellungen von Einstein sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet.
Ich möchte im Folgenden—im Sinne der axiomatischen Methode—aus drei einfachen Axiomen ein neues System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der gestellten Probleme enthalten ist. Die genauere Ausführung sowie vor allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor." (Proofs, 1)

> intimate, hitherto hidden processes in the interior of atoms will receive an explanation; and in particular that generally a reduction of all physical constants to mathematical constants must be possible—whereby the possibility approaches that physics in principle becomes a science of the type of geometry: surely the highest glory of the axiomatic method, which, as we have seen, here takes into its service the powerful instruments of analysis, namely the calculus of variations and the theory of invariants.[88]

Hilbert's final remarks about the status of his theory *vis à vis* Einstein's work on gravitation strikingly parallel Minkowski's assessment of the relation of his four-dimensional formulation to Einstein's special theory; not just providing a mathematical framework for existing results, but developing a genuinely novel physical theory, which, properly understood, turns out to be a part of mathematics.[89]

Fig. 1 provides a graphical survey of the deductive structure of Hilbert's theory. The main elements listed above are connected by arrows; mathematical implications are represented by straight arrows and inferences based on heuristic reasoning by curved arrows. As the figure shows, apart from the field equations, Hilbert's results can be divided into two fairly distinct clusters: one comprises the implications for electromagnetism (right-hand side of the diagram); the other, the implications for the understanding of energy conservation (left-hand side of the diagram). While the assertions concerning energy conservation are not essential for deriving the other results, they depend on practically all the other parts of this theory. The main link between the two clusters is clearly Theorem I. Although no assertion of Hilbert's theory is derived directly from Theorem I, it motivates both the relation between energy conservation and coordinate restrictions and the link between electromagnetism and gravitation.

The analysis of the deductive structure of Hilbert's theory thus confirms that Theorem I is indeed the *Leitmotiv* of the theory. The two clusters of results obviously are also related to what he considered the two main physical touchstones of his theory: Mie's theory of electromagnetism and energy conservation. On the other hand, neither Newton's theory of gravitation nor any other parts of mechanics are mentioned by Hilbert. Einstein's imprint on Hilbert's theory was more of a mathematical or structural nature than a physical one.

---

88 "Wie man sieht, genügen bei sinngemäßer Deutung die wenigen einfachen in den Axiomen I, II, III ausgesprochenen Annahmen zum Aufbau der Theorie: durch dieselbe werden nicht nur unsere Vorstellungen über Raum, Zeit und Bewegung von Grund aus in dem von Einstein geforderten Sinne umgestaltet, sondern ich bin auch der Überzeugung, daß durch die hier aufgestellten Grundgleichungen die intimsten, bisher verborgenen Vorgänge innerhalb des Atoms Aufklärung erhalten werden und insbesondere allgemein eine Zurückführung aller physikalischen Konstanten auf mathematische Konstanten möglich sein muß—wie denn überhaupt damit die Möglichkeit naherückt, daß aus der Physik im Prinzip eine Wissenschaft von der Art der Geometrie werde: gewiß der herrlichste Ruhm der axiomatischen Methode, die hier wie wir sehen die mächtigen Instrumente der Analysis nämlich, Variationsrechnung und Invariantentheorie, in ihre Dienste nimmt." (Proofs, 13)

89 For Minkowski, see (Walter 1999).

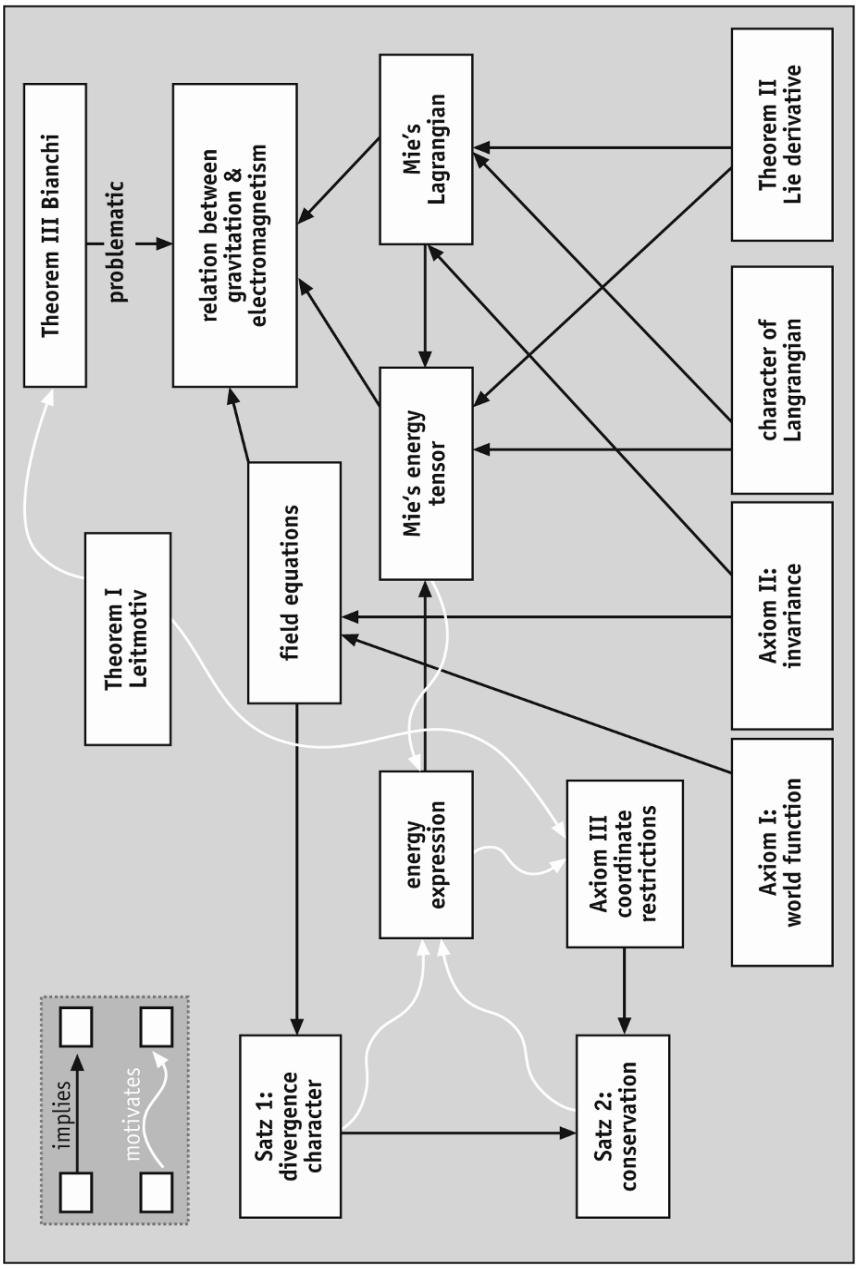

Figure 1: Deductive Structure of the Proofs (1915)

4. HILBERT'S PHYSICS AND EINSTEIN'S MATHEMATICS: THE EXCHANGE OF LATE 1915

4.1 What Einstein Could Learn From Hilbert

The Hilbert-Einstein correspondence begins with Einstein's letter of 7 November 1915.[90] That November was the month during which Einstein's theory of gravitation underwent several dramatic changes documented by four papers he presented to the Prussian Academy, culminating in the definitive version of the field equations in the paper submitted 25 November.[91] On 4 November Einstein submitted his first note, in which he abandoned the *Entwurf* field equations and replaced them with equations derived from the Riemann tensor (Einstein 1915a); he included the proofs of this paper in his letter to Hilbert. In spite of this radical modification of the field equations, the structure of Einstein's theory remained essentially unchanged from that of the non-covariant 1913 *Entwurf* theory. In both, the requirement of energy-momentum conservation is linked to a restriction to adapted coordinate systems. In Einstein's 4 November paper, this restriction implies the following equation (Einstein 1915a, 785):

$$\sum_{\alpha\beta}\frac{\partial}{\partial x_\alpha}\left(g^{\alpha\beta}\frac{\partial lg\sqrt{-g}}{\partial x_\beta}\right) = -\kappa\sum_{\sigma}T_\sigma^\sigma. \tag{87}$$

Einstein pointed out one immediate consequence for the choice of an adapted coordinate system:

> Equation (21a) [i.e. (87)] shows the impossibility of so choosing the coordinate system that $\sqrt{-g}$ equals 1, because the scalar of the energy tensor cannot be set to zero.[92]

That the scalar [i.e. the trace] of the energy-momentum tensor cannot vanish is obvious if one takes Einstein's standard example (a swarm of non-interacting particles or incoherent "dust") as the source of the gravitational field: the trace of its energy-momentum tensor equals the mass density of the dust. However, the physical meaning of condition (87) was entirely obscure. It was therefore incumbent upon Einstein to find a physical interpretation of it or to modify his theory once more in order to get rid of it. He soon succeeded in doing both, and formulated his new view in an addendum to the first note, published on 11 November (Einstein 1915b).

On 12 November 1915 he reported his success to Hilbert:

> For the time being, I just thank you cordially for your kind letter. Meanwhile, the problem has made new progress. Namely, it is possible to compel *general* covariance by means of the postulate $\sqrt{-g} = 1$; Riemann's tensor then furnishes the gravitational

90 Einstein to David Hilbert, 7 November 1915, (CPAE 8, 191).

91 See (Einstein 1915e).

92 "Aus Gleichung (21a) [i.e. (87)] geht hervor, daß es unmöglich ist, das Koordinatensystem so zu wählen, daß $\sqrt{-g}$ gleich 1 wird; denn der Skalar des Energietensors kann nicht zu null gemacht werden." (Einstein 1915a, 785)

> equations directly. If my present modification (which does not change the equations) is legitimate, then gravitation must play a fundamental role in the structure of matter. My own curiosity is impeding my work![93]

What had happened? Einstein had noticed that the condition $\sum_{\sigma} T_{\sigma}^{\sigma} = 0$, which follows from setting $\sqrt{-g} = 1$ in (87), can be related to an electromagnetic theory of matter: in Maxwell's theory, the vanishing of its trace is a characteristic property of the electromagnetic energy-momentum tensor. Thus, if one assumes all matter to be of electromagnetic origin, the vanishing of its trace becomes a fundamental property of the energy-momentum tensor. This has two important consequences: Condition (87) is no longer an inexplicable restriction on the admissible coordinate systems, and the 4 November field equations can be seen as a particular form of generally-covariant field equations based on the Ricci tensor. From the perspective of the 11 November revision, the condition $\sqrt{-g} = 1$ turns out to be nothing more than an arbitrary but convenient choice of coordinate systems.

The core of Einstein's new theory is strikingly simple. The left-hand side of the gravitational field equations is now simply the Ricci tensor and the right-hand side an energy-momentum tensor, the trace of which has to vanish:[94]

$$R_{\mu\nu} = -\kappa T_{\mu\nu} \qquad \sum_{\sigma} T_{\sigma}^{\sigma} = 0. \tag{88}$$

What distinguishes these field equations from the final equations presented on 25 November is an additional term on the right-hand side of the equations involving the trace of the energy-momentum tensor, which now need not vanish:[95]

$$R_{\mu\nu} = -\kappa\left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right). \tag{89}$$

Remarkably enough, in the winter of 1912/13 Einstein had considered the linearized form of these field equations, but discarded them because they were not compatible with his expectation of how the Newtonian limit should result.[96] He had also then considered and rejected field equations of the form (88), just because they imply

93 "Ich danke einstweilen herzlich für Ihren freundlichen Brief. [Das] Problem hat unterdessen einen neuen Fortschritt gemacht. Es lässt sich nämlich durch das Postulat $\sqrt{-g} = 1$ die *allgemeine* Kovarianz erzwingen; der Riemann'sche Tensor liefert dann direkt die Gravitationsgleichungen. Wenn meine jetzige Modifikation (die die Gleichungen nicht ändert) berechtigt ist, dann muss die Gravitation im Aufbau der Materie eine fundamentale Rolle spielen. Die Neugier erschwert mir die Arbeit!" Einstein to David Hilbert, 12 November 1915, (CPAE 8, 194).

94 See (Einstein 1915b, 801 and 800).

95 See (Einstein 1915e, 845).

96 See Doc. 10 of (CPAE 4), "Pathways out of Classical Physics …", "Einstein's Zurich Notebook", (both in vol. 1 of this series), and the "Commentary" (in vol. 2).

the condition $\sum_{\sigma} T_{\sigma}^{\sigma} = 0$. At that time, this condition seemed unacceptable because the trace of the energy-momentum tensor of ordinary matter does *not* vanish.

The prehistory of Einstein's 11 November paper thus confronts us with a puzzle: Why did he consider it to be such a decisive advance beyond his 4 November paper and not just a possible alternative interpretation of his previous results; and why did he now so readily accept the trace-condition $\sum_{\sigma} T_{\sigma}^{\sigma} = 0$ that earlier had led him to reject this very theory? What impelled Einstein's change of perspective in November 1915?

The answer seems to lie in the changed context, within which Einstein formulated his new approach: in particular, his interaction with Hilbert. As will become evident, it would have been quite uncharacteristic of him to adopt the new approach so readily had it not been for current discussions of the electrodynamic worldview and his feeling that he was now in competition with Hilbert.[97]

In his addendum, Einstein directly referred to the supporters of the electrodynamic worldview:

> One now has to remember that, in accord with our knowledge, "matter" is not to be conceived as something primitively given, or physically simple. There even are those, and not just a few, who hope to be able to reduce matter to purely electrodynamic processes, which of course would have to be done in a theory more complete than Maxwell's electrodynamics.[98]

Only this context explains Einstein's highly speculative and fragmentary comments on an electromagnetic model of matter. That, in November 1915, Einstein conceived of a field theory of matter as a goal in its own right is also supported by his correspondence, which makes it clear that this perspective was shaped by his rivalry with Hilbert. We have already cited Einstein's letter to Hilbert, in which he wrote:

> If my present modification (which does not change the equations) is legitimate, then gravitation must play a fundamental role in the structure of matter. My own curiosity is impeding my work![99]

And when, in a letter of 14 November, Hilbert claimed to have achieved the unification of gravitation and electromagnetism, Einstein responded:

---

97 For a discussion of Hilbert's reaction to what he must have seen as an intrusion by Einstein into his domain, see (Sauer 1999, 542–543).

98 "Es ist nun daran zu erinnern, daß nach unseren Kenntnissen die "Materie" nicht als ein primitiv Gegebenes, physikalisch Einfaches aufzufassen ist. Es gibt sogar nicht wenige, die hoffen, die Materie auf rein elektromagnetische Vorgänge reduzieren zu können, die allerdings einer gegenüber Maxwells Elektrodynamik vervollständigten Theorie gemäß vor sich gehen würden." (Einstein 1915b, 799)

99 "Wenn meine jetzige Modifikation (die die Gleichungen nicht ändert) berechtigt ist, dann muss die Gravitation im Aufbau der Materie eine fundamentale Rolle spielen. Die Neugier erschwert mir die Arbeit!" Einstein to David Hilbert, 12 November 1915, (CPAE 8, 194).

> Your investigation interests me tremendously, especially since I often racked my brain to construct a bridge between gravitation and electromagnetics.[100]

A few days later (after calculating the perihelion shift on the basis of the new theory), he expressed himself similarly:

> In these last months I had great success in my work. *Generally covariant* gravitation equations. *Perihelion motions explained quantitatively.* The role of gravitation in the structure of matter. You will be astonished. I worked dreadfully hard; it is remarkable that one can sustain it.[101]

When one examines Einstein's previous writings on gravitation, published and unpublished, one finds no trace of an attempt to unify gravitation and electromagnetism. He had never advocated the electromagnetic worldview. On the contrary, he was apparently disinterested in Mie's attempt at a unification of gravitation and electrodynamics, not finding it worth mentioning in his 1913 review of contemporary gravitation theories.[102]

And soon after completion of the final version of general relativity, Einstein reverted to his earlier view that general relativity could make no assertions about the structure of matter:

> From what I know of Hilbert's theory, it makes use of an assumption about electrodynamic processes that—apart from the treatment of the gravitational field—is closely connected to Mie's. Such a specialized approach is not in accordance with the point of view of general relativity. The latter actually only provides the gravitational field law, and quite unambiguously so when general covariance is required.[103]

Einstein's mid-November 1915 pursuit of a relation between gravitation and electromagnetism was, then, merely a short-lived episode in his search for a relativistic theory of gravitation. Its novelty is confirmed by a footnote in the addendum:

> In writing the earlier paper, I had not yet realized that the hypothesis $\sum T^{\mu}_{\mu} = 0$ is, in principle, admissible.[104]

---

100 "Ihre Untersuchung interessiert mich gewaltig, zumal ich mir schon oft das Gehirn zermartert habe, um eine Brücke zwischen Gravitation und Elektromagnetik zu schlagen." Einstein to David Hilbert, 15 November 1915, (CPAE 8, 199).

101 "Ich habe mit grossem Erfolg gearbeitet in diesen Monaten. *Allgemein kovariante* Gravitationsgleichungen. *Perihelbewegungen quantitativ erklärt*. Rolle der Gravitation im Bau der Materie. Du wirst staunen. Gearbeitet habe ich schauderhaft angestrengt; sonderbar, dass man es aushält." Einstein to Michele Besso, 17 November 1915, (CPAE 8, 201).

102 See (Einstein 1913).

103 "Soviel ich von Hilbert's Theorie weiss, bedient sie sich eines Ansatzes für das elektrodynamische Geschehen, der sich [— a]bgesehen von der Behandlung des Gravitationsfeldes — eng an Mie anschliesst. Ein derartiger spezieller Ansatz lässt sich aus dem Gesichtspunkte der allgemeinen Relativität nicht begründen. Letzterer liefert eigentlich nur das Gesetz des Gravitationsfeldes, und zwar ganz eindeutig, wenn man allgemeine Kovarianz fordert." Einstein to Arnold Sommerfeld, 9 December 1915, (CPAE 8, 216).

104 "Bei Niederschrift der früheren Mitteilung war mir die prinzipielle Zulässigkeit der Hypothese $\sum T^{\mu}_{\mu} = 0$ noch nicht zu Bewußtsein gekommen." (Einstein 1915b, 800)

It thus seems quite clear that Einstein's temporary adherence to an electromagnetic theory of matter was triggered by Hilbert's work, which he attempted to use in order to solve a problem that had arisen in his own theory, and that he dropped it when he solved this problem in a different way.

So this whole episode might appear to be a bizarre and unnecessary detour. A closer analysis of the last steps of Einstein's path to general relativity shows, however, that the solution depended crucially on this detour, and hence indirectly on Hilbert's work. In fact, Einstein successfully calculated the perihelion shift of Mercury on the basis of his 11 November theory.[105] The condition $\sqrt{-g} = 1$, implied by the assumption of an electromagnetic origin of matter (see (87)), was essential for this calculation, which Einstein considered a striking confirmation of his audacious hypothesis on the constitution of matter, definitely favoring this theory over that of 4 November.[106] The 11 November theory also turned out to be the basis for a new understanding of the Newtonian limit, which allowed Einstein to accept the field equations of general relativity as the definitive solution to the problem of gravitation. Ironically, Hilbert's most important contribution to general relativity may have been enhancing the credibility of a speculative and ultimately untenable physical hypothesis that guided Einstein's final mathematical steps towards the completion of his theory.

Einstein submitted his perihelion paper on 18 November 1915. In a footnote, appended after its completion, Einstein observed that, in fact, the hypothesis of an electromagnetic origin of matter is unnecessary for the perihelion shift calculation. He announced a further modification of his field equations, finally reaching the definitive version of his theory.[107] On the same day, Einstein wrote to Hilbert, acknowledging receipt of Hilbert's work, including a system of field equations:

> The system [of field equations] you give agrees—as far as I can see—exactly with that which I found in the last few weeks and have presented to the Academy.[108]

---

105 See (Einstein 1915c).

106 See (Einstein 1915d): the abstract of this paper, probably by Einstein, summarizes the issue: "Es wird gezeigt, daß die allgemeine Relativitätstheorie die von Leverrier entdeckte Perihelbewegung des Merkurs qualitativ und quantitativ erklärt. Dadurch wird die Hypothese vom Verschwinden des Skalars des Energietensors der "Materie" bestätigt. Ferner wird gezeigt, daß die Untersuchung der Lichtstrahlenkrümmung durch das Gravitationsfeld ebenfalls eine Möglichkeit der Prüfung dieser wichtigen Hypothese bietet." ("It will be shown that the theory of general relativity explains qualitatively and quantitatively the perihelion motion of Mercury, which was discovered by Leverrier. Thus the hypothesis of the vanishing of the scalar of the energy tensor of "matter" is confirmed. Furthermore, it is shown that the analysis of the bending of light by the gravitational field also offers a way of testing this important hypothesis.")

107 See (Einstein 1915c, 831).

108 "Das von Ihnen gegebene System [of field equations] stimmt - soweit ich sehe - genau mit dem überein, was ich in den letzten Wochen gefunden und der Akademie überreicht habe." Einstein to David Hilbert, 18 November 1915, (CPAE 8, 201–202). For discussion of what Einstein may have received from Hilbert, see below.

Einstein emphasized that the real difficulty had not been the formulation of generally-covariant field equations, but in showing their agreement with a physical requirement: the existence of the Newtonian limit. Stressing his priority, he mentioned that he had considered such equations three years earlier:

> … it was hard to recognize that these equations form a generalisation, and indeed a simple and natural generalisation, of Newton's law. It has just been in the last few weeks that I succeeded in this (I sent you my first communication), whereas 3 years ago with my friend Grossmann I had already taken into consideration the only possible generally covariant equations, which have now been shown to be the correct ones. We had only heavy-heartedly distanced ourselves from it, because it seemed to me that the physical discussion had shown their incompatibility with Newton's law.[109]

Einstein's statement not only characterized his own approach, but indirectly clarified his ambivalent position with regard to Hilbert's theory. While evidently fascinated by the perspective of unifying gravitation and electromagnetism, he now recognized that, at least in Hilbert's case, this involved the risk of neglecting the sound foundation of the new theory of gravitation in the classical theory.

4.2 What Hilbert Could Learn from Einstein

Hilbert must have seen Einstein's letter of 12 November, announcing publication of new insights into a fundamental role of gravitation in the constitution of matter, as a threat to his priority.[110] At any rate, Hilbert hastened public presentation of his results. His response of 13 November gave a brief sketch of his theory and announced a 16th November seminar on it:

> Actually, I wanted first to think of a quite palpable application for physicists, namely valid relations between physical constants, before obliging with my axiomatic solution to your great problem. But since you are so interested, I would like to develop my th[eory] in very complete detail on the coming Tuesday, that is, the day after the day after tomorrow (the 16th of this mo.). I find it ideally beautiful math[ematically], and also insofar as calculations that are not completely transparent do not occur at all, and absolutely compelling in accordance with the axiom[atic] meth[od] and therefore rely on its reality. As a result of a gen. math. theorem, the (generalized Maxwellian) electrody. eqs. appear as a math. consequence of the gravitation eqs., so that gravitation and electrodynamics are actually not at all different. Furthermore, my energy concept forms the basis: $E = \Sigma(e_s t^s + e_{ih} t^{ih})$, [the t^s corresponds to p^s in Hilbert's papers, etc.] which is likewise a general invariant [see (56)], and from this then also follow from a very simple

109 "schwer war es, zu erkennen, dass diese Gleichungen eine Verallgemeinerung, und zwar eine einfache und natürliche Verallgemeinerung des Newton'schen Gesetzes bilden. Dies gelang mir erst in den letzten Wochen (meine erste Mitteilung habe ich Ihnen geschickt), während ich die einzig möglichen allgemein kovarianten Gleichungen, [die] sich jetzt als die richtigen erweisen, schon vor 3 Jahren mit meinem Freunde Grossmann in Erwägung gezogen hatte. Nur schweren Herzens trennten wir uns davon, weil mir die physikalische Diskussion scheinbar ihre Unvereinbarkeit mit Newtons Gesetz ergeben hatte."

110 This aspect of the Hilbert-Einstein relationship was first discussed in (Sauer 1999), where the chronology of events is carefully reconstructed.

> axiom the 4 still-missing "spacetime equations" $e_s = 0$. I derived most pleasure in the discovery already discussed with Sommerfeld that the usual electrical energy results when a certain absolute invariant is differentiated with respect to the gravitation potentials and then g are set = 0,1.[111]

This letter presents the essential elements of Hilbert's theory as presented in the Proofs. His reference to "the missing spacetime equations" suggests that he saw these equations and their relation to the energy concept as an issue common to his theory and Einstein's.

Einstein responded on 15 November 1915, declining the invitation to come to Göttingen on grounds of health.[112] Instead, he asked Hilbert for the proofs of his paper. As mentioned above, by 18 November Hilbert had fulfilled Einstein's request. He could not have sent the typeset Proofs, which are dated 6 December, so he must have sent a manuscript on 20 November, presumably corresponding to his talk. Since the Proofs are also dated 20 November, this manuscript may well have presented practically the same version of his theory. On 19 November, a day after Einstein announced his successful perihelion calculation to Hilbert, the latter sent his congratulations, making clear once more that the physical problems facing Hilbert's theory were of a rather different nature:

> Many thanks for your postcard and cordial congratulations on conquering perihelion motion. If I could calculate as rapidly as you, in my equations the electron would correspondingly have to capitulate, and simultaneously the hydrogen atom would have to produce its note of apology about why it does not radiate.
> I would be grateful if you were to continue to keep me up-to-date on your latest advances.[113]

---

111 "Ich wollte eigentlich erst nur für die Physiker eine ganz handgreifliche Anwendung nämlich treue Beziehungen zwischen den physikalischen Konstanten überlegen, ehe ich meine axiomatische Lösung ihres grossen Problems zum Besten gebe. Da Sie aber so interessiert sind, so möchte ich am kommenden Dienstag also über-über morgen (d. 16 d. M.) meine Th. ganz ausführlich entwickeln. Ich halte sie für math. ideal schön auch insofern, als Rechnungen, die nicht ganz durchsichtig sind, garnicht vorkommen. und absolut zwingend nach axiom. Meth., und baue deshalb auf ihre Wirklichkeit. In Folge eines allgem. math. Satzes erscheinen die elektrody. Gl. (verallgemeinerte Maxwellsche) als math. Folge der Gravitationsgl., so dass Gravitation u. Elektrodynamik eigentlich garnichts verschiedenes sind. Desweiteren bildet mein Energiebegriff die Grundlage: $E = \Sigma(e_s t^s + e_{ih} t^{ih})$, die ebenfalls eine allgemeine Invariante ist, und daraus folgen dann aus einem sehr einfachen Axiom die noch fehlenden 4 "Raum-Zeitgleichungen" $e_s = 0$. Hauptvergnügen war für mich die schon mit Sommerfeld besprochene Entdeckung, dass die gewöhnliche elektrische Energie herauskommt, wenn man eine gewisse absolute Invariante mit den Gravitationspotentialen differenziert und dann $g = 0, 1$ setzt." David Hilbert to Einstein, 13 November 1915, (CPAE 8, 195).

112 Einstein to David Hilbert, 15 November 1915, (CPAE 8, 199).

113 "Vielen Dank für Ihre Karte und herzlichste Gratulation zu der Ueberwältigung der Perihelbewegung. Wenn ich so rasch rechnen könnte, wie Sie, müsste bei meinen Gleichg entsprechend das Elektron kapituliren und zugleich das Wasserstoffatom sein Entschuldigungszettel aufzeigen, warum es nicht strahlt. Ich werde Ihnen auch ferner dankbar sein, wenn Sie mich über Ihre neuesten Fortschritte auf dem Laufenden halten." David Hilbert to Einstein, 19 November 1915, (CPAE 8, 202).

No doubt Einstein fulfilled this request to keep Hilbert up to date. His definitive paper on the field equations, submitted 25 November and published 2 December, must have been on Hilbert's desk within a day or two. In contrast to all earlier versions of his theory, Einstein now showed that energy-momentum conservation does not imply additional coordinate restrictions on the field equations (89). He also made clear that these field equations fulfill the requirement of having a Newtonian limit and allow derivation of the perihelion shift of Mercury.

Our analysis of the Proofs suggests that neither the astronomical implications of Einstein's theory nor the latter's treatment of the Newtonian limit directly affected Hilbert's theory since they lay outside its scope, as Hilbert then perceived it. But Einstein's insight that energy-momentum conservation does not lead to a restriction on admissible coordinate systems was of crucial significance for Hilbert. As we have seen, in Hilbert's theory the entire complex of results on energy-momentum conservation was structured by a logic paralleling that of Einstein's earlier non-covariant theory. Moreover, Theorem I, Hilbert's *Leitmotiv*, was motivated by Einstein's hole argument that generally-covariant field equations cannot have unique solutions. His definitive paper of 25 November did not explicitly mention the hole argument, but simply took it for granted that his new generally-covariant field equations avoid such difficulties.[114] Hilbert may well have checked that Einstein's definitive field equations were actually compatible[115] with the equations that follow from Hilbert's variational principle, which he had not explicitly calculated—or at least not included in the Proofs, and this compatibility would certainly have been reassuring for Hilbert. But the fact that the hole argument evidently no longer troubled Einstein must have led Hilbert to question his *Leitmotiv*, with its double role of motivating coordinate restrictions and providing the link between gravitation and electromagnetism.

Thus, Einstein's paper of 25 November 1915 represented a major challenge for Hilbert's theory. As we shall see when discussing the published version of Hilbert's paper, while Einstein temporarily took over Hilbert's physical perspective, Hilbert appears to have accepted the mathematical implications of Einstein's rejection of the hole argument.

4.3 Cooperation in the Form of Competition

In a situation such as we have described, in which the interaction between two people working on closely related problems changes the way in which each of them proceeds, it is not easy for the individuals to assess their own contributions. While Einstein was happy to have found in Hilbert one of the few colleagues, if not the only one, who appreciated and understood the nature of his work on gravitation, he also

114 The fact that these equations were supported by Einstein's successful calculation of the perihelion shift made it impossible for Hilbert simply to disregard them.

115 Compatible, but not the same, because of the trace term, and because of the different treatment of the stress-energy tensor, as discussed elsewhere in this paper.

resented the way in which Hilbert took over some of his results without, as Einstein saw it, giving him due credit. Einstein wrote to his friend Heinrich Zangger on 26 November 1915 with regard to his newly-completed theory:

> The theory is beautiful beyond comparison. However, only *one* colleague has really understood it, and he is seeking to "partake" [*nostrifizieren*] in it (Abraham's expression) in a clever way. In my personal experience I have hardly come to know the wretchedness of mankind better than as a result of this theory and everything connected to it. But it does not bother me.[116]

Einstein's reaction becomes particularly understandable in the light of his prior positive experience of collaboration with his friend, the mathematician Marcel Grossmann. Grossmann had restricted himself to putting his superior mathematical competence at Einstein's service.[117] What Hilbert offered was not cooperation but competition. Hilbert may well have been upset by Einstein's anticipation in print, in his paper of 11 November, of what Hilbert felt to be his idea of a close link between gravitation and the structure of matter. Even more disturbing may have been the fact that, contrary to Hilbert's assertion in the Proofs, Einstein's final formulation of his theory required no restriction on general covariance. But it is not clear exactly when Hilbert abandoned all non-covariant elements of his program, in particular his approach to the energy problem and consequent restriction to a preferred class of coordinate systems.[118]

Hilbert evidently learned of Einstein's resentment over lack of recognition by Hilbert, possibly as a result of Einstein's letter of 18 November pointing out his priority in setting up generally-covariant field equations. In any case, he began to introduce changes in his Proofs on or after 6 December, documented by handwritten marginalia, changes which not only acknowledge Einstein's priority but attempt to placate him. Hilbert's revision also provides an indication of the content of Einstein's complaints. He revised the programmatic statement in the introduction of his paper (his insertion is rendered in italics):

> In the following — in the sense of the axiomatic method — I would like to develop, *essentially* from three simple axioms a ~~new~~ system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented.[119]

---

116 "Die Theorie ist von unvergleichlicher Schönheit. Aber nur *ein* Kollege hat sie wirklich verstanden und der eine sucht sie auf geschickte Weise zu "*nostrifizieren*" (Abraham'scher Ausdruck). Ich habe in meinen persönlichen Erfahrungen kaum je die Jämmerlichkeit der Menschen besser kennen gelernt wie gelegentlich dieser Theorie und was damit zusammenhängt. Es ficht mich aber nicht an." Einstein to Heinrich Zangger, 26 November 1915, (CPAE 8, 205). See the discussion of "nostrification" above.

117 See the editorial note "Einstein on Gravitation and Relativity: The Collaboration with Marcel Grossmann" in (CPAE 4, 294–301).

118 According to (Sauer 1999, 562), Hilbert had found the new energy expression by 25 January 1916.

119 "Ich möchte im Folgenden - im Sinne der axiomatischen Methode - *wesentlich* aus drei einfachen Axiomen ein ~~neues~~ System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der gestellten Probleme enthalten ist." (Proofs, 1)

The insertion "wesentlich" was presumably motivated by Hilbert's recognition that his theory actually presupposed additional assumptions of substantial content, such as the assumption of a split of the Lagrangian into gravitational and electromagnetic parts and the assumption that the latter does not depend on derivatives of the metric (see section 3). A further assumption was the requirement that the gravitational part of the Lagrangian not involve derivatives of the metric higher than second order. Einstein had justified this requirement by the necessity for the theory to have a Newtonian limit, and it may have been Einstein's argument that drew Hilbert's attention to the fact that his theory was actually based on a much wider array of assumptions than his axiomatic presentation had indicated. More remarkably, in characterizing his system of equations, Hilbert deleted the word "neu," a clear indication that he had read Einstein's 25 November paper and recognized that the equations implied by his own variational principle are formally equivalent (because of where the trace term occurs) to Einstein's if Hilbert's electrodynamic stress-energy tensor is substituted for the unspecified one on the right-hand side of Einstein's field equations.

Hilbert's next change was presumably related to a complaint by Einstein about the lack of proper acknowledgement for what he considered to be one of his fundamental contributions, the introduction of the metric tensor as the mathematical representation of the gravitational potentials. Hilbert had indeed given the impression that Einstein's merit was confined to asking the right questions, while Hilbert provided the answers.

Hilbert's revised description of these gravitational potentials reads (his insertion is again rendered in italics):

> The quantities characterizing the events at w_s shall be:
>
> 1) The ten gravitational potentials *first introduced by* Einstein, $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ having the character of a symmetric tensor with respect to arbitrary transformation of the world parameter w_s;
>
> 2) The four electrodynamic potentials q_s having the character of a vector in the same sense.[120]

The next change represents an even more far-going recognition that Hilbert could not simply claim the results in his paper as parts of "his theory," as if it had nothing substantial in common with that of Einstein:

> The guiding motive for setting up ~~my~~ *the* theory is given by the following theorem, the proof of which I will present elsewhere.[121]

---

120 "Die das Geschehen in w_s charakterisierenden Größen seien:
1) die zehn *von Einstein zuerst eingeführten* Gravitationspotentiale $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ mit symmetrischem Tensorcharakter gegenüber einer beliebigen Transformation der Weltparameter w_s;
2) die vier elektrodynamischen Potentiale q_s mit Vektorcharakter im selben Sinne." (Proofs, 1)

121 "Das Leitmotiv für den Aufbau ~~meiner~~ *der* Theorie liefert der folgende mathematische Satz, dessen Beweis ich an einer anderen Stelle darlegen werde." (Proofs, 2)

Hilbert's final marginal notation consists of just an exclamation mark next to a minor correction of the energy expression (39)—perhaps evidence that he had identified this expression as the central problem in the Proofs. While Hilbert's first annotations were presumably intended as revisions of a text that was going to remain basically unchanged, this exclamation mark signals the abandonment of such an attempt at revision. At this point, perhaps it dawned upon Hilbert that Einstein's results forced him to rethink his entire approach.

Hilbert's recognition of the problematic character of his treatment of energy-momentum conservation appears to have been solely in reaction to Einstein's results and not as a consequence of any internal dynamics (see section 3) of the development of his theory.[122] Indeed, as our analysis of the deductive structure of Hilbert's theory showed, this treatment is well anchored in the remainder of his theory without in turn having much effect on the remainder. Hence, there was no "internal friction" that could have driven a further development of Hilbert's theory. On the contrary, since the link between energy-momentum conservation and coordinate restrictions was motivated by Hilbert's Theorem I, Einstein's abandonment of this link left Hilbert at a loss, as we have argued above. But the way in which energy-momentum conservation was connected to other results of his theory also suggested how to modify it in the direction indicated by Einstein: Hilbert had to find a new energy expression that does not imply a coordinate restriction but is still connected with Mie's energy-momentum tensor. Precisely the decoupling of his energy expression from the physical consequences of Hilbert's theory made such a modification possible. Hilbert gave up immediate publication and began to rework his theory. By early 1916 had he arrived at results that made possible this rewriting of his paper and its submission for publication; by mid-February 1916, Paper 1, which we will discuss in the following section, was in press.[123]

Meanwhile, having emerged triumphant from the exchange of November 1915, Einstein offered a reconciliation to Hilbert:

> There has been a certain ill-feeling between us, the cause of which I do not want to analyze. I have struggled against the feeling of bitterness attached to it, and this with complete success. I think of you again with unmarred friendliness and ask you to try to do the same with me. Objectively it is a shame when two real fellows who have extricated themselves somewhat from this shabby world do not afford each other mutual pleasure.[124]

122 For a different view, see (Sauer 1999, 570).

123 For a detailed chronology, see the reconstruction in (Sauer 1999, 560–565).

124 "Es ist zwischen uns eine gewisse Verstimmung gewesen, deren Ursache ich nicht analysieren will. Gegen das damit verbundene Gefühl der Bitterkeit habe ich gekämpft, und zwar mit vollständigem Erfolge. Ich gedenke Ihrer wieder in ungetrübter Freundlichkeit, und bitte Sie, dasselbe bei mir zu versuchen. Es ist objektiv schade, wenn sich zwei wirkliche Kerle, die sich aus dieser schäbigen Welt etwas herausgearbeitet haben, nicht gegenseitig zur Freude gereichen." Einstein to David Hilbert, 20 December 1915, (CPAE 8, 222). The "schäbige [.] Welt" probably refers to World War I—given Einstein and Hilbert's critical attitude to the war.

5. HILBERT'S ASSIMILATION OF EINSTEIN'S RESULTS: THE THREE PUBLISHED VERSIONS OF HIS FIRST PAPER

5.1 The New Energy Concept—An Intermediary Solution

As we have seen, modification of Hilbert's treatment of energy-momentum conservation was the most urgent step necessitated by Einstein's results of 25 November 1915. First of all, the energy-momentum conservation law should not involve coordinate restrictions but be an invariant equation. Second, the modified energy expression should still involve Mie's energy-momentum tensor; otherwise the link between gravitation and electromagnetism, fundamental to Hilbert's program, would be endangered. Third, to accord with Hilbert's understanding of energy-momentum conservation, the new energy concept must still satisfy a divergence equation. As we shall show, Hilbert's modification of his energy expression was guided by these criteria, but its relation to a physical interpretation remained as tenuous as ever.[125] The next section concerns the effect of the new energy concept on the deductive structure of Hilbert's theory.

In the introductory discussion of energy, Paper 1 emphasizes that only axioms I and II are required:

> The most important aim is now the formulation of the concept of energy, and the derivation of the energy theorem solely on the basis of the two axioms I and II.[126]

This emphasis is in contrast with the treatment in the Proofs, in which the energy concept is closely related to axiom III, which was dropped in Paper 1. Hilbert then proceeds exactly as in the Proofs, introducing a polarization of the Lagrangian with respect to the gravitational variables (see the definition of P_g, (20)):

$$P_g(\sqrt{g}H) = \sum_{\mu,\nu,k,l} \left(\frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} p^{\mu\nu} + \frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}_k} p^{\mu\nu}_k + \frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}_{kl}} p^{\mu\nu}_{kl} \right). \tag{90}$$

In contrast to (37), however, Hilbert polarizes $\sqrt{g}H$ instead of H. Clearly, his aim was to formulate an equation analogous to (45), but with only a divergence term on the right-hand side. Indeed, since:

$$P(\sqrt{g}H) = \sqrt{g}PH + H\sum_{\mu,\nu} \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} p^{\mu\nu}, \tag{91}$$

use of $\sqrt{g}H$ eliminates the first term of the right-hand side of (45), giving:

---

125 For a discussion of Hilbert's concept of energy, see also (Sauer 1999, 548–550), which stresses the mathematical roots of this concept.

126 "Das wichtigste Ziel ist nunmehr die Aufstellung des Begriffes der Energie und die Herleitung des Energiesatzes allein auf Grund der beiden Axiome I und II." (Hilbert 1916, 400)

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g}(a^l + b^l)}{\partial w_l} = \sum_{\mu,\nu} [\sqrt{g}H]_{\mu\nu} p^{\mu\nu}. \tag{92}$$

Since the right-hand side vanishes due to the field equations, this equation is of just the desired form.

The way in which Hilbert obtained (92) closely parallels that used in the Proofs, i.e. by splitting off divergence terms. He starts out by noting that:

$$a^l = \sum_{\mu,\nu,k} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} A^{\mu\nu}_k, \tag{93}$$

where $A^{\mu\nu}_k$ is the covariant derivative of $p^{\mu\nu}$, is a contravariant vector.

Then he observes that:

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g} a^l}{\partial w_l} \tag{94}$$

no longer contains the second derivatives of $p^{\mu\nu}$, and hence can be written:

$$\sqrt{g} \sum_{\mu,\nu,k} (B_{\mu\nu} p^{\mu\nu} + B^k_{\mu\nu} p^{\mu\nu}_k), \tag{95}$$

where $B^k_{\mu\nu}$ is a tensor. Finally, Hilbert forms the vector:

$$b^l = \sum_{\mu,\nu} B^l_{\mu\nu} p^{\mu\nu}, \tag{96}$$

obtaining (92).

He next forms the expression for the electromagnetic variables analogous to (92) (see the definition of P_q, (20) above):

$$P_q(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g} c^l}{dw_l} = \sum_k [\sqrt{g}H]_k p_k, \tag{97}$$

with:

$$c^l = \sum_k \frac{\partial H}{\partial q_{kl}} p_k. \tag{98}$$

Adding (92) and (97), and taking account of the field equations, Hilbert could thus write:

$$P(\sqrt{g}H) = \sum_l \frac{\partial \sqrt{g}(a^l + b^l + c^l)}{\partial w_l}. \tag{99}$$

The final step consists in also rewriting the left-hand side of this equation as a divergence, using (91), which is expanded as:

$$P(\sqrt{g}H) \;=\; \sqrt{g}PH + H\sum_s \left(\frac{\partial\sqrt{g}}{\partial w_s}p^s + \sqrt{g}\,p^s_s\right); \tag{100}$$

using Theorem II (see (22)),[127] he then obtained:

$$P(\sqrt{g}H) \;=\; \sqrt{g}\sum_z \frac{\partial H}{\partial w_s}p^s + H\sum_s \left(\frac{\partial\sqrt{g}}{\partial w_s}p^s + \sqrt{g}\,p^s_s\right) \;=\; \sum_s \frac{\partial\sqrt{g}Hp^s}{\partial w_s}, \tag{101}$$

and, in view of (99),

$$\sum_l \frac{\partial}{\partial w_l}\sqrt{g}(Hp^l - a^l - b^l - c^l) \;=\; 0. \tag{102}$$

This equation could have been interpreted as giving the energy expression since, being an invariant divergence, it satisfies two of the three criteria mentioned above. But it is not related to Mie's energy-momentum tensor. So Hilbert adds yet another term $-d^l$ to the expression in the parenthesis in (102):

$$d^l \;=\; \frac{1}{2\sqrt{g}}\sum_{k,\,s}\frac{\partial}{\partial w_k}\left\{\left(\frac{\partial\sqrt{g}H}{\partial q_{lk}} - \frac{\partial\sqrt{g}H}{\partial q_{kl}}\right)p^s q_s\right\}, \tag{103}$$

which does not alter its character since d^l is a contravariant vector (because:

$$\frac{\partial H}{\partial q_{lk}} - \frac{\partial H}{\partial q_{kl}} \tag{104}$$

is an antisymmetric tensor) that satisfies the identity:

$$\sum_l \frac{\partial\sqrt{g}d^l}{\partial w_l} \;=\; 0. \tag{105}$$

Hilbert concluded:

> Let us now define
>
> $$e^l \;=\; Hp^l - a^l - b^l - c^l - d^l \tag{106}$$
>
> as the *energy vector, then the energy vector is a contravariant vector, which moreover depends linearly on the arbitrarily chosen vector* p^s, *and satisfies identically for that choice of this vector* p^s *the invariant energy equation*
>
> $$\sum_l \frac{\partial\sqrt{g}e^l}{\partial w_l} \;=\; 0.\,^{128} \tag{107}$$

127 In Paper 1, this is the only purpose for which this form of Theorem II is explicitly introduced. However, (23) presumably already had been derived from it.

While Hilbert did not explicitly introduce the condition that his energy vector be related to Mie's energy-momentum tensor, it seems to be the guiding principle of his calculation. Apparently, he wanted this connection to appear to be the result of an independently-justified definition of this vector.

In effect, starting from (106) and taking into account definitions (98) and (103), Hilbert obtained for the contribution to the energy originating from the electromagnetic term L in the Lagrangian:

$$Lp^l - \sum_k \frac{\partial L}{\partial q_{kl}} p_k - \frac{1}{2\sqrt{g}} \sum_{k,s} \frac{\partial}{\partial w_k} \left\{ \left(\frac{\partial \sqrt{g} L}{\partial q_{lk}} - \frac{\partial \sqrt{g} L}{\partial q_{kl}} \right) p^s q_s \right\}. \tag{108}$$

Using the field equations and (27), this can be rewritten as:

$$\sum_{s,k} \left(L\delta_s^l - \frac{\partial L}{\partial M_{lk}} M_{sk} - \frac{\partial L}{\partial q_l} q_s \right) p^s, \tag{109}$$

which corresponds to the right-hand side of (36), the generally-covariant generalization of Mie's electromagnetic energy-momentum tensor, contracted with p^s.

In contradistinction to the Proofs, Theorem II and (36) no longer explicitly enter this demonstration. Theorem II enters implicitly by determining the form in which the electromagnetic variables enter the Lagrangian (see (27)). Hilbert still needed Theorem II to derive his "first result," that is, to show that this energy-momentum can be written as the variational derivative of $\sqrt{g}L$ with respect to the gravitational potentials. Furthermore, (36) allows Hilbert to argue that, due to the field equations (see (72)), the electromagnetic energy and energy-vector e^l can be expressed exclusively in terms of K, the gravitational part of the Lagrangian; so that they depend only on the metric tensor and not on the electromagnetic potentials and their derivatives. Whereas, in the Proofs, this result had been an immediate consequence of the definition of the energy and of the field equations (see (49)), now it follows only with the help of Theorem II.

While Hilbert had succeeded in satisfying his heuristic criteria as well as the new challenge of deriving an invariant energy equation, the status of this equation within his theory had become more precarious. An analysis of the deductive structure of Hilbert's theory in Paper 1 (see Fig. 2) shows that it still comprises two main clusters of results: those concerning the implications of gravitation for electromagnetism and those concerning energy conservation. But the latter cluster is now even more isolated

128 "Definieren wir nunmehr [(106); (14) in the original text] als den *Energievektor, so ist der Energievektor ein kontravarianter Vektor, der noch von dem willkürlichen Vektor* p^s *linear abhängt und identisch für jene Wahl dieses Vektors* p^s *die invariante Energiegleichung* [(107)] *erfüllt*." (Hilbert 1916, 402)

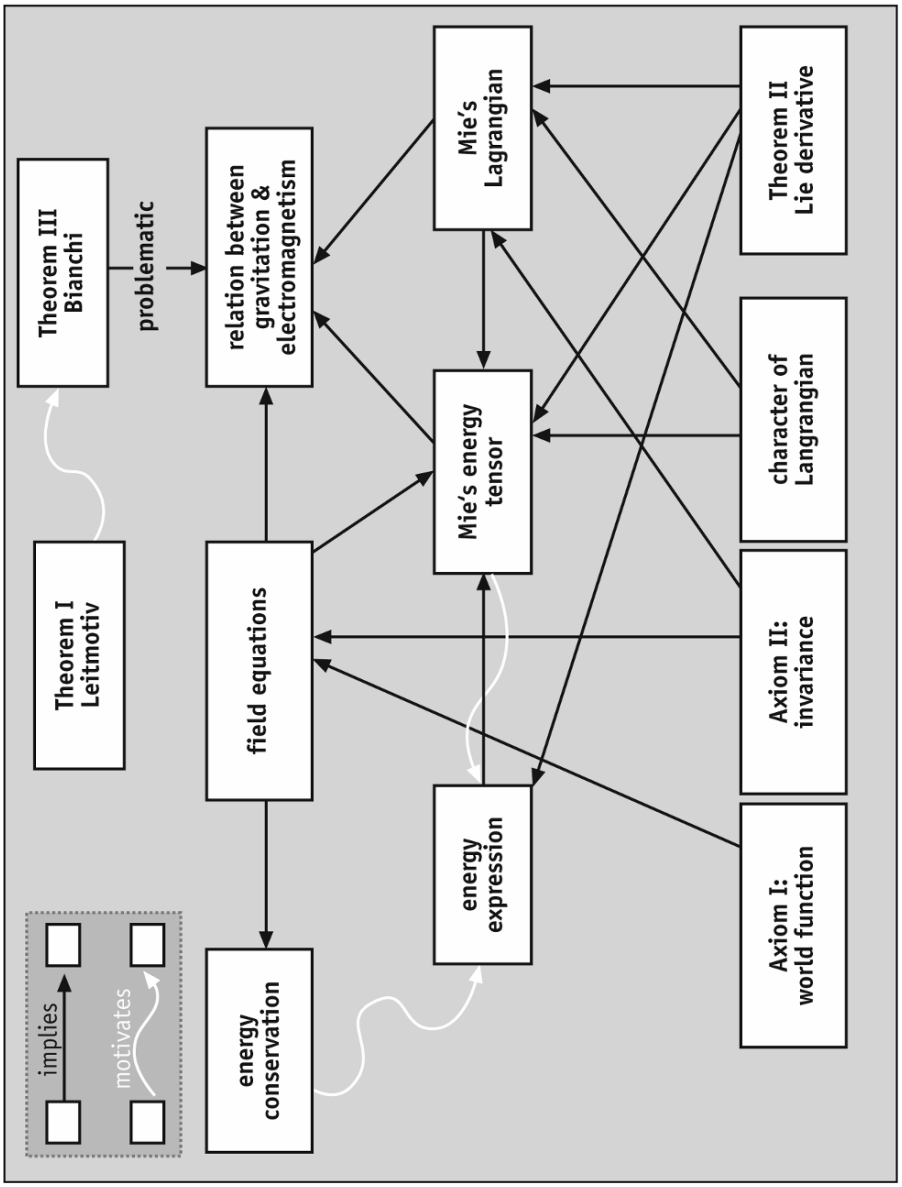

Figure 2: Deductive Structure of Paper 1 (1916)

from the rest of his theory than in the Proofs. Indeed, the new energy concept is no longer motivated by Hilbert's powerful Theorem I, but only by arguments concerning the formal properties of energy-momentum conservation and the link with Mie's energy-momentum tensor. It plays no role in deriving any other results of Hilbert's theory, nor does it serve to integrate this theory with other physical theories, a key function of the energy concept since its formulation in the 19th century. Therefore, it is not surprising that this concept only played a transitional role and was eventually replaced by the understanding of energy-momentum conservation developed by Einstein, Klein, Noether, and others.[129]

In fact, neither the physical significance nor the mathematical status of Hilbert's new energy concept was entirely clear. Physically Hilbert had failed to show that his energy equation (107) gave rise to a familiar expression for energy-momentum conservation in the special-relativistic limit, or to demonstrate that his equation was compatible with the form of energy-momentum conservation in a gravitational field that Einstein had established in 1913 (see (11)). Eventually, Felix Klein succeeded in clarifying the relation between Hilbert's and Einstein's expressions. He decomposed (107) into 140 equations and showed that 136 of these actually have nothing to do with energy-momentum conservation, while the remaining 4 correspond to those given by Einstein.[130] Mathematically, in 1917 Emmy Noether and Felix Klein found that equation (107) actually is an identity, and not a consequence of the field equations, as is the case for conservation equations in classical physics.[131] Similar identities follow for the Lagrangian of any generally-covariant variational problem. As a consequence, Hilbert's counting of equations no longer works: he assumed that his variational principle gives rise to 10 gravitational field equations plus 4 identities, which he identified with the electromagnetic equations; and that energy-momentum conservation is represented by additional equations, originally linked to coordinate restrictions. Einstein's abandonment of coordinate restrictions together with the deeper investigation of energy-momentum conservation by Noether, Klein, Einstein, and others, confronted Hilbert's approach with a severe challenge: They questioned the organization of his theory into two more-or-less independent domains, energy-momentum conservation and the implications of gravitation for electromagnetism. We shall argue that Hilbert responded to this challenge by further adapting his theory to the framework provided by general relativity.

5.2 Hilbert's Reorganization of His Theory in Paper 1

The challenge presented by Einstein's abandonment of coordinate restrictions and adoption of generally-covariant field equations forced Hilbert to reorganize his the-

129 For discussion, see (Rowe, 1999).

130 See (Klein 1918a, 179–185).

131 See (Klein 1917; 1918a) and also (Noether 1918). For a thorough discussion of the contemporary research on energy-momentum conservation, see (Rowe, 1999).

ory. As we have seen, he had to demonstrate the compatibility between his variational principle and Einstein's field equations (from which he had succeeded strikingly in deriving Mercury's perihelion shift), and completely rework his treatment of energy conservation. Hilbert treated both issues at the end of Paper 1. Energy conservation was no longer tied to Theorem I and its heuristic consequences as in the Proofs, but was treated along with other results of Hilbert's theory. The structure of Paper 1 is thus:[132]

1. Basic Framework (Hilbert 1916, 395–398)

   Axioms I and II, Theorem I, and the combined field equations of gravitation and electromagnetism for an arbitrary Lagrangian

2. Basic Theorems (Hilbert 1916, 398–400)

   Theorems II and III

3. New Energy Expression and Derivation of the New Energy Equation (Hilbert 1915, 400–402)

4. Implications for the Relation between Electromagnetism and Gravitation (Hilbert 1915, 402–407)

   the split of the Lagrangian into gravitational and the electrodynamical terms, the form of Mie's Lagrangian, its relation to his energy tensor, the explicit form of the gravitational field equations, and the relation between electromagnetic and gravitational field equations.

Apart from the technical and structural revisions necessitated by the new energy expression, practically all other changes concern the relation of his theory to Einstein's. Throughout Paper 1, Hilbert followed the tendency, already manifest in the marginal additions to the Proofs, to put greater emphasis on Einstein's contributions while maintaining his claim to have developed an independent approach. In the opening paragraph, Hilbert changed the order in which he mentioned Mie and Einstein. In the Proofs he wrote:

> The far reaching ideas and the formation of novel concepts by means of which Mie constructs his electrodynamics, and the prodigious problems raised by Einstein, as well as his ingeniously conceived methods of solution, have opened new paths for the investigation into the foundations of physics.[133]

In Paper 1 we read instead:

> The vast problems posed by Einstein as well as his ingeniously conceived methods of solution, and the far-reaching ideas and formation of novel concepts by means of which

132 For a sketch of Hilbert's revisions of Paper 1, see also (Corry 1999a, 517–522).

133 "Die tiefgreifenden Gedanken und originellen Begriffsbildungen vermöge derer Mie seine Elektrodynamik aufbaut, und die gewaltigen Problemstellungen von Einstein sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet." (Proofs, 1)

> Mie constructs his electrodynamics, have opened new paths for the investigation into the foundations of physics.[134]

A footnote lists all of Einstein's publications on general relativity starting with his major 1914 review, and including the definitive paper submitted on 25 November. Although this makes clear that Hilbert must have revised his paper after that date, he failed to change the dateline of his contribution (as did Felix Klein and Emmy Noether in their contributions to the discussion of Hilbert's work in the same journal[135]). It remained "Vorgelegt in der Sitzung vom 20. November 1915," which creates the erroneous impression that there were no subsequent substantial changes in Paper 1.

The next sentence, while combining this claim with a more explicit recognition of what he considered the achievements of his predecessors, shows that Hilbert had not renounced his claim to having solved the problems posed by Mie and Einstein. In the corrected Proofs this sentence reads:

> In the following—in the sense of the axiomatic method — I would like to develop, /*essentially* from three simple axioms a ~~new~~ system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented.[136]

In Paper 1, it reads:

> In the following — in the sense of the axiomatic method — I would like to develop, essentially from two simple axioms, a new system of basic equations of physics, of ideal beauty and containing, I believe, *simultaneously* the solution to the problems of Einstein and of Mie. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.[137]

Although in a marginal note in the proofs version he had changed "his theory" to "the theory," he now returned to the original version:

> The guiding motive for constructing my theory is provided by the following theorem, the proof of which I shall present elsewhere.[138]

---

134 "Die gewaltigen Problemstellungen von Einstein sowie dessen scharfsinnige zu ihrer Lösung ersonnenen Methoden und die tiefgreifenden Gedanken und originellen Begriffsbildungen vermöge derer Mie seine Elektrodynamik aufbaut, haben der Untersuchung über die Grundlagen der Physik neue Wege eröffnet." (Hilbert 1916, 395)

135 See (Klein 1918a; Noether 1918).

136 "Ich möchte im Folgenden — im Sinne der axiomatischen Methode —/*wesentlich* aus drei einfachen Axiomen ein ~~neues~~ System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der gestellten Probleme enthalten ist." (Proofs, 1)

137 "Ich möchte im Folgenden - im Sinne der axiomatischen Methode - wesentlich aus zwei einfachen Axiomen ein neues System von Grundgleichungen der Physik aufstellen, die von idealer Schönheit sind, und in denen, wie ich glaube, die Lösung der Probleme von Einstein und Mie gleichzeitig enthalten ist. Die genauere Ausführung sowie vor Allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor." (Hilbert 1916, 395)

138 "Das Leitmotiv für den Aufbau meiner Theorie liefert der folgende mathematische Satz, dessen Beweis ich an einer anderen Stelle darlegen werde." (Hilbert 1916, 396)

Although Hilbert had earlier argued that his *Leitmotiv* suggested the need for four additional non-covariant equations to ensure a unique solution, he now dropped all mention of the subject of coordinate restrictions. He simply did not address the question of why, in spite of Einstein's hole argument against this possibility, it is possible to use generally-covariant field equations unsupplemented by coordinate restrictions. The only remnant in Paper 1 of the entire problem is his newly-introduced designation of the world-parameters as "allgemeinste Raum-Zeit-Koordinaten."

The significant result that Hilbert's variational principle gives rise to gravitational field equations formally equivalent to those of Einstein's 25 November theory is rather hidden in Hilbert's presentation, only appearing as an intermediate step in his demonstration that the electromagnetic field equations are a consequence of the gravitational ones. The newly-introduced passage reads:

> Using the notation introduced earlier for the variational derivatives with respect to the $g^{\mu\nu}$, the gravitational equations, because of (20) [i.e. (16)], take the form
>
> $$[\sqrt{g}K]_{\mu\nu} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{110}$$
>
> The first term on the left hand side becomes
>
> $$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\right), \tag{111}$$
>
> as follows easily without calculation from the fact that $K_{\mu\nu}$, apart from $g_{\mu\nu}$, is the only tensor of second rank and K the only invariant, that can be formed using only the $g^{\mu\nu}$ and their first and second differential quotients, $g^{\mu\nu}_k$, $g^{\mu\nu}_{kl}$.
>
> The resulting differential equations of gravitation appear to me to be in agreement with the grand concept of the theory of general relativity established by Einstein in his later treatises.[139]

Hilbert's argument for avoiding explicit calculation of $[\sqrt{g}K]_{\mu\nu}$, which he later withdrew (see below), is indeed untenable; there are many invariants and tensors of second rank that can be constructed from the Riemann tensor. Even if one further requires such tensors and invariants to be linear in the Riemann tensor, the crucial coefficient of the trace term still remains undetermined. The explicit form of the field equations given in Paper 1 and not found in the Proofs, appears to be a direct response to Einstein's publication of 25 November; but a footnote appended to this

---

139 "Unter Verwendung der vorhin eingeführten Bezeichungsweise für die Variationsableitungen bezüglich der $g^{\mu\nu}$ erhalten die Gravitationsgleichungen wegen (20) [i.e. (16)] die Gestalt [(110); (21) in the original text]. Das erste Glied linker Hand wird [(111)] wie leicht ohne Rechnung aus der Tatsache folgt, daß $K_{\mu\nu}$ außer $g_{\mu\nu}$ der einzige Tensor zweiter Ordnung und K die einzige Invariante ist, die nur mit den $g^{\mu\nu}$ und deren ersten und zweiten Differentialquotienten $g^{\mu\nu}_k, g^{\mu\nu}_{kl}$ gebildet werden kann.

Die so zu Stande kommenden Differentialgleichungen der Gravitation sind, wie mir scheint, mit der von Einstein in seinen späteren Abhandlungen aufgestellten großzügigen Theorie der allgemeinen Relativität im Einklang." (Hilbert 1916, 404–405)

passage gives a generic reference to all four of Einstein's 1915 Academy publications. His cautious reference to the apparent agreement between his results and Einstein's, presumably motivated by their different frameworks, adds to the impression that Hilbert actually arrived independently at the explicit form of the gravitational field equations.

The concluding paragraph of Paper 1 acknowledges Hilbert's debt to Einstein in a more indirect way. The beginning of this paragraph of the Proofs had given the impression that Einstein posed the problems while Hilbert offered the solutions:

> As one can see, the few simple assumptions expressed in axioms I, II, III suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense called for by Einstein ...[140]

In Paper 1, Hilbert deleted the reference to axiom III and replaced "in dem von Einstein geforderten Sinne" by "in dem von Einstein dargelegten Sinne":

> As one can see, the few simple assumptions expressed in axioms I and II suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense explained by Einstein ...[141]

5.3 Einstein's Energy in Hilbert's 1924 Theory

In 1924 Hilbert published revised versions of Papers 1 and 2 (Hilbert 1924).[142] Meanwhile important developments had taken place, such as the rapid progress of quantum physics, which changed the scientific context of Hilbert's results. But it was undoubtedly the further clarifications of the significance of energy-momentum conservation in general relativity, already mentioned in the preceding sections, that affected his theory most directly. In correspondence between Hilbert and Klein (published in part in 1918),[143] this topic played a central role without, however, leading to an explicit reformulation of Hilbert's theory. Without going into detail about this important strand in the history of general relativity, we shall focus on its effect on Hilbert's 1924 revisions. In spite of the reassertion of his goal of providing foundations for all of physics, his theory was, in effect, transformed into a variation on the themes of general relativity.

---

140 "Wie man sieht, genügen bei sinngemäßer Deutung die wenigen einfachen in den Axiomen I, II, III ausgesprochenen Annahmen zum Aufbau der Theorie: durch dieselbe werden nicht nur unsere Vorstellungen über Raum, Zeit und Bewegung von Grund aus in dem von Einstein geforderten Sinne umgestaltet ..." (Proofs, 13).

141 "Wie man sieht, genügen bei sinngemäßer Deutung die wenigen einfachen in den Axiomen I und II ausgesprochenen Annahmen zum Aufbau der Theorie: durch dieselbe werden nicht nur unsere Vorstellungen über Raum, Zeit und Bewegung von Grund aus in dem von Einstein dargelegten Sinne umgestaltet ..." (Hilbert 1916, 407).

142 In the following, we will refer to the 1924 revision of Paper 1 as "Part 1" and to that of Paper 2 as "Part 2," designations which correspond to Hilbert's own division of his 1924 paper into "Teil 1" (pp. 2–11) and "Teil 2" (pp. 11–32).

143 See (Klein 1917).

On a purely technical level, Hilbert's revisions of Paper 1 appear to be rather modest; the most important one concerns Theorem III (the contracted Bianchi identities), now labelled Theorem 2. Following a suggestion by Klein (Klein 1917, 471–472), Hilbert extended this theorem to include the electromagnetic variables:

> Theorem 2. Let J, as in Theorem 1, be an invariant depending on $g^{\mu\nu}$, $g^{\mu\nu}_l$, $g^{\mu\nu}_{lk}$, q_s, q_{sk}; and as above, let $[\sqrt{g}J]_{\mu\nu}$ denote the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$, and $[\sqrt{g}J]_\mu$, the variational derivative with respect to q_μ. Introduce, furthermore, the abbreviations [(112)]:
>
> $$\begin{aligned} i_s &= \sum_{\mu,\nu}([\sqrt{g}J]_{\mu\nu}g^{\mu\nu}_s + [\sqrt{g}J]_\mu q_{\mu s}), \\ i^l_s &= -2\sum_\mu [\sqrt{g}J]_{\mu s}g^{\mu l} + [\sqrt{g}J]_l q_s, \end{aligned} \tag{112}$$
>
> then the [following] identities hold
>
> $$i_s = \sum_l \frac{\partial i^l_s}{\partial x_l} \qquad (s = 1, 2, 3, 4).^{144} \tag{113}$$

He revised its proof accordingly.

A second, small, but significant change concerns the gravitational field equations. Hilbert now tacitly withdrew his previous claim that no derivation was needed, instead sketching a derivation and writing them, like Einstein, with the energy-momentum tensor as source. As in the earlier versions, he derived (72) but now in the form:[145]

$$[\sqrt{g}K]_{\mu\nu} = -\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}. \tag{114}$$

After writing down the electromagnetic field equations, Hilbert proceeded to sketch the following evaluation of the terms in (114):

> To determine the expression for $[\sqrt{g}K]_{\mu\nu}$, first specialize the coordinate system so that at the world point under consideration all the $g^{\mu\nu}_s$ vanish. In this way one finds:
>
> $$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}g_{\mu\nu}K\right). \tag{115}$$
>
> If, for the tensor
>
> $$-\frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} \tag{116}$$
>
> we introduce the symbol $T_{\mu\nu}$, then the gravitational field equations can be written as

144 "Theorem 2. Wenn J, wie im Theorem 1, eine von $g^{\mu\nu}$, $g^{\mu\nu}_l$, $g^{\mu\nu}_{lk}$, q_s, q_{sk} abhängige Invariante ist, und, wie oben, die Variationsableitungen von $\sqrt{g}J$ bez. $g^{\mu\nu}$ mit $[\sqrt{g}J]_{\mu\nu}$, bez. q_μ mit $[\sqrt{g}J]_\mu$ bezeichnet werden, und wenn ferner zur Abkürzung: [(112)] gesetzt wird, so gelten die Identitäten [(113); (7) in the original text]." (Hilbert 1924, 5)

145 See (Hilbert 1924, 7).

$$K_{\mu\nu} - \frac{1}{2} g_{\mu\nu} K = T_{\mu\nu}.^{146} \tag{117}$$

Although the introduction of Einstein's notation for the energy-momentum tensor may appear as no more than an adaptation of Hilbert's notation to the by-then standard usage, it actually effected a major revision in the structure of his theory. The energy-momentum tensor became the central knot binding together the physical implications of Hilbert's theory.

First of all, it served, as Hilbert's energy expressions had previously done, to relate the derivative of Mie's Lagrangian (see (34) or (36)) to Mie's energy-momentum tensor. But, in contrast to Paper 1, Mie's energy-momentum tensor no longer served as a criterion for choosing the energy-expression. The new energy expression, which Hilbert now took over from Einstein, was supported by much more than just this single result. It had emerged from the development of special-relativistic continuum physics by Minkowski, Abraham, Planck, Laue,[147] and others; and been validated by numerous applications to various areas of physics, including general relativity.

By introducing the equation:

$$T_{\mu\nu} = -\frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}}, \tag{118}$$

Hilbert had returned, in a sense, to the approach of the Proofs, establishing a relation between the energy concept and the derivative of the electromagnetic Lagrangian (see (49)). He still did not make clear that this relation does not single out Mie's theory, but actually holds more generally. Introducing the notations:

$$\frac{\partial L}{\partial q_{sk}} = \frac{\partial L}{\partial M_{ks}} = H^{ks}, \tag{119}$$

and:

$$\frac{\partial L}{\partial q_k} = r^k, \tag{120}$$

As in the proofs version, Hilbert again used (35), which he now rewrites as:

$$-\frac{2}{\sqrt{g}} \sum_{\mu} \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} = L \delta^m_\nu - \sum_s H^{ms} M_{\nu s} - r^m q_\nu, \tag{121}$$

146 "Um den Ausdruck von $[\sqrt{g}K]_{\mu\nu}$ zu bestimmen, spezialisiere man zunächst das Koordinatensystem so, daß für den betrachteten Weltpunkt die $g^{\mu\nu}_s$ sämtlich verschwinden. Man findet auf diese Weise: [(115)]. Führen wir noch für den Tensor [(116)] die Bezeichnung $T_{\mu\nu}$ ein, so lauten die Gravitationsgleichungen [(117)]." See (Hilbert 1924, 7–8).

147 For the first systematic development of relativistic continuum mechanics, see (Laue 1911a; 1911b). For further discussion, see Einstein's "Manuscript on the Special Theory of Relativity" (CPAE 4, Doc. 1, 91–98; Janssen and Mecklenburg 2006).

(see (36)). On the basis of this equation, Hilbert claims, in almost exactly the same words as in the earlier versions, that there is a necessary connection between the theories of Mie and Einstein:

> Hence the [following] representation of $T_{\mu\nu}$ results:
>
> $$T_{\mu\nu} = \sum_{\mu} g_{\mu m} T_{\nu}^{m}$$
> $$T_{\nu}^{m} = \frac{1}{2}\left\{ L\delta_{\nu}^{m} - \sum_{s} H^{ms} M_{\nu s} - r^{m} q_{\nu} \right\}. \tag{122}$$
>
> The expression on the right agrees with Mie's electromagnetic energy tensor, and thus we find that Mie's electromagnetic energy tensor is nothing but the generally-invariant tensor resulting from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ — a circumstance which gave me the first hint of the necessary close connection between Einstein's theory of general relativity and Mie's electrodynamics, and which convinced me of the correctness of the theory developed here.[148]

While Hilbert's claim remained unchanged, what he had done actually was to specialize the source term left arbitrary in Einstein's field equations. The nature of this source term can be specified on the level of the Lagrangian or of the energy-momentum tensor, and these two ways are obviously equivalent if a Lagrangian exists—but this relation is in no way peculiar to Mie's theory. The fact that the energy expression in Paper 1 was specifically chosen to produce Mie's energy-momentum tensor had obscured this circumstance, now made rather obvious by the introduction of Einstein's arbitrary energy-momentum tensor. It was no doubt difficult for Hilbert to draw this conclusion because it contradicted his program, according to which electromagnetism should arise as an effect of gravitation.

The situation was similar for Hilbert's second important application of Einstein's energy-momentum tensor, the derivation of a relation between the gravitational and electromagnetic field equations. After recognition of the close relation between the contracted Bianchi identities and energy-momentum conservation in general relativity, it was necessary for Hilbert to reconsider the link he believed he had established between the two groups of field equations. Energy-momentum conservation now played a central role in his approach, turning the link between gravitation and electromagnetism into a mere by-product. It existed, not because of any deep intrinsic connection between these two areas of physics, but due to the introduction of electromagnetic potentials into the variational principle. With the same logic, one

148 "Demnach ergibt sich für $T_{\mu\nu}$ die Darstellung: [(122)]. Der Ausdruck rechts stimmt überein mit dem Mie'schen elektromagnetischen Energietensor, und wir finden also, daß der Mie'sche elektromagnetische Energietensor ist nichts anderes als der durch Differentiation der Invariante L nach den Gravitationspotentialen $g^{\mu\nu}$ entstehende allgemein invariante Tensor—ein Umstand, der mich zum ersten Mal auf den notwendigen engen Zusammenhang zwischen der Einsteinschen allgemeinen Relativitätstheorie und der Mie'schen Elektrodynamik hingewiesen und mir die Überzeugung von der Richtigkeit der hier entwickelten Theorie gegeben hat." (Hilbert 1924, 9)

could argue that *any* form of matter giving rise to a stress-energy tensor derivable from a Lagrangian involving the metric tensor is an effect of gravitation.

This weakened link is reflected in Hilbert's new way of obtaining the desired link between gravitation and electromagnetism. Following Klein's suggestion, in Part 1 Hilbert treated the contracted Bianchi identities in parallel for both the gravitational and the electromagnetic terms in the Lagrangian:

> The application of Theorem 2 to the invariant K yields:
>
> $$\sum_{\mu\nu}[\sqrt{g}K]_{\mu\nu}g_s^{\ \mu\nu} + 2\sum_m \frac{\partial}{\partial x_m}\Big(\sum_\mu [\sqrt{g}K]_{\mu s}g^{\mu m}\Big) = 0. \tag{123}$$
>
> Its application to L yields:[149]
>
> $$\begin{aligned}&\sum_{\mu\nu}(-\sqrt{g}T_{\mu\nu})g_s^{\ \mu\nu} + 2\sum_m \frac{\partial}{\partial x_m}(-\sqrt{g}T_s^{\ m})\\ &+\sum_\mu [\sqrt{g}L]_\mu q_{\mu s} - \sum_\mu \frac{\partial}{\partial x_\mu}([\sqrt{g}L]_\mu q_s) = 0 \qquad (s = 1, 2, 3, 4).\end{aligned} \tag{124}$$

Previously, he had derived only the first set of identities and made use of them in order to derive (83). Now Hilbert showed that both sets of identities yield the equations for energy-momentum conservation that had been central to Einstein's work since 1912. Following the work of Einstein and others, Hilbert also made clear that these equations are related to the equations of motion for the sources of the stress-energy tensor,[150] and represent a generalization of energy-momentum conservation laws in special relativity:

> As a consequence of the basic equations of electrodynamics, we obtain from this:
>
> $$\sum_{\mu\nu}\sqrt{g}T_{\mu\nu}g_s^{\ \mu\nu} + 2\sum_m \frac{\partial}{\partial x_m}\sqrt{g}T_s^{\ m} = 0. \tag{125}$$
>
> These equations also result as a consequence of the gravitational equations due to (15a) [i.e. (123)]. Their interpretation is that they are the basic equations of mechanics. In the case of special relativity, when the $g_{\mu\nu}$ are constants, they reduce to the equations
>
> $$\sum \frac{\partial T_s^{\ m}}{\partial x_m} = 0, \tag{126}$$
>
> which express the conservation of energy and momentum.[151]

149 "Die Anwendung des Theorems 2 auf die Invariante K liefert: [(123); (15a) in the original text.] Die Anwendung auf L ergibt: [(124); (15b) in the original text.]" (Hilbert 1924, 9–10)

150 See (Havas 1989, Klein 1917; 1918a; 1918b).

151 "Als Folge der elektrodynamischen Grundgleichungen erhalten wir hieraus: [(125); (16) in the original text.] Diese Gleichungen ergeben sich auch als Folge der Gravitationsgleichungen, auf Grund von (15a) [i.e. (123)]. Sie haben die Bedeutung der mechanischen Grundgleichungen. Im Falle der speziellen Relativität, wenn die $g_{\mu\nu}$ Konstante sind, gehen sie über in die Gleichungen [(126)] welche die Erhaltung von Energie und Impuls ausdrücken." (Hilbert 1924, 10)

Hilbert thus anchored his theory in the same physical foundation that had provided Einstein's search for general relativity with a stable point of reference. Only after having done this did Hilbert turn to his original goal, the link between gravitation and electromagnetism, the problematic character of which we have discussed above:

> From the identities (15b) [i.e. (124)], there follow from the equations (16) [i.e. (125)]:
>
> $$\sum_{\mu}[\sqrt{g}L]_{\mu}q_{\mu s} - \sum_{\mu}\frac{\partial}{\partial x_{\mu}}([\sqrt{g}L]_{\mu}q_{s}) = 0 \tag{127}$$
>
> or
>
> $$\sum_{\mu}\left\{M_{\mu s}[\sqrt{g}L]_{\mu} + q_{s}\frac{\partial}{\partial x_{\mu}}[\sqrt{g}L]_{\mu}\right\} = 0; \tag{128}$$
>
> i.e., four independent linear relations between the basic equations of electrodynamics (5) and their first derivations follow from the gravitational equations (4). This is the precise mathematical expression of the connection between gravitation and electrodynamics, which dominates the entire theory.[152]

The deductive structure of Part 1 shows the fundamental changes with respect to Paper 1 (see Fig. 3) and the central role of Einstein's energy-momentum tensor in this reorganization. In fact, this tensor suggested the particular form in which Hilbert rewrote the gravitational field equations, established the link between gravitation and electromagnetism (in terms of the choice of a specific source), and, of course, was fundamental to Hilbert's new formulation of energy-momentum conservation.

This revised deductive structure has a kernel, consisting of the variational principle, field equations, and energy-momentum conservation, that is—both from a formal and a physical perspective —fully equivalent to the kernel of Einstein's formulation of general relativity. Clearly, Hilbert's deductive presentation places greater emphasis on a variational principle than does Einstein; and the mathematically more elegant formulation of the variational principle, based on the Ricci scalar, contributes to this emphasis. Therefore, this variational formulation of general relativity is today rightly associated with Hilbert's name. On the other hand, Hilbert's original aim, the derivation of electromagnetism as an effect of gravitation, plays only a marginal role in Part 1 and still suffers from the problems indicated above. The links between the main components that had substantiated Hilbert's claim of a special relation between Mie's theory and Einstein's have been weakened, being held together only by the choice of a specific source. This link is thus no longer central to an approach presenting an alternative to that of Einstein, being little more than an attempt to supplement

152 "Aus den Gleichungen (16) [i.e. (125)] folgt auf Grund der Identitäten (15b) [i.e. (124)]: [(127)] oder [(128); (17) in the original text] d.h. aus den Gravitationsgleichungen (4) folgen vier voneinander unabhängige lineare Relationen zwischen den elektrodynamischen Grundgleichungen (5) und ihren ersten Ableitungen. Dies ist der genaue mathematische Ausdruck für den Zusammenhang zwischen Gravitation und Elektrodynamik, der die ganze Theorie beherrscht." See the comments on (83), (Hilbert 1924, 10).

Einstein's general framework with a specific physical content, Mie's electrodynamics—an attempt that is now based on the firm foundations of general relativity.

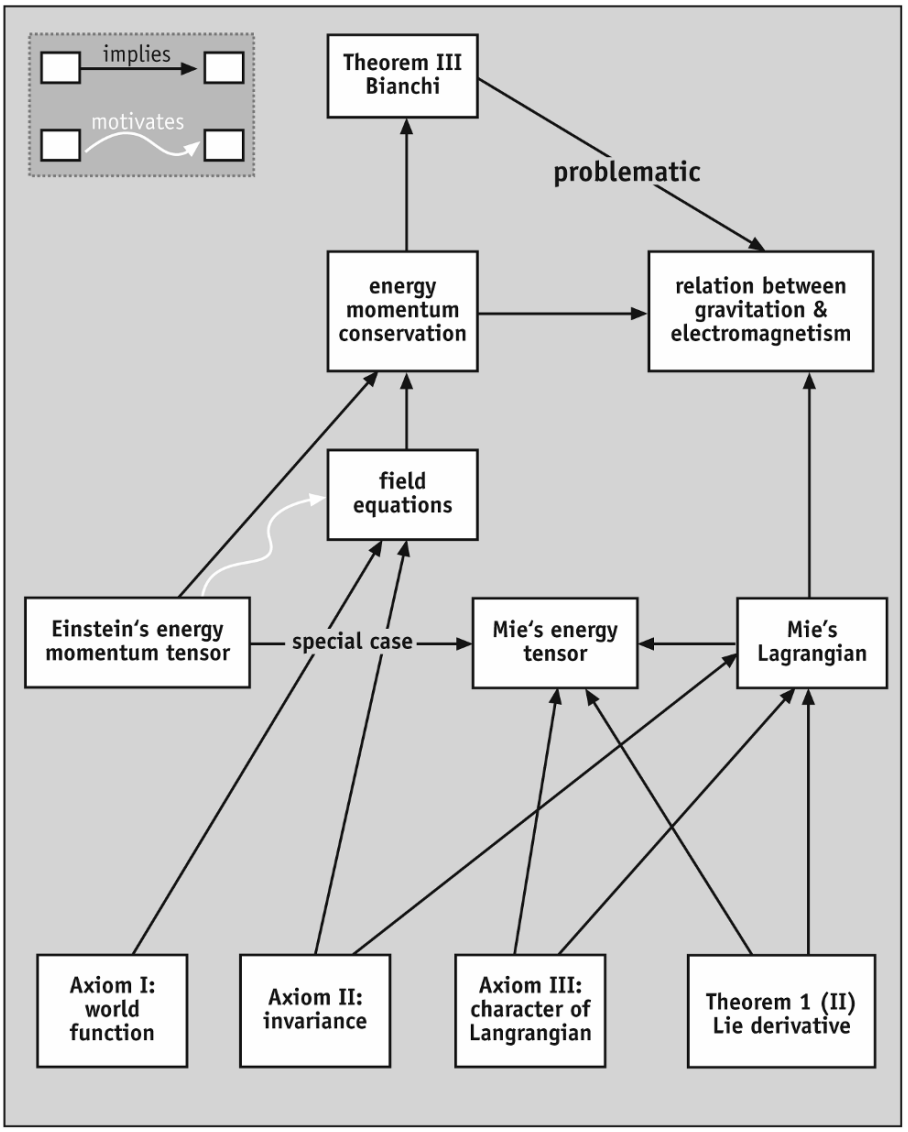

Figure 3: Deductive Structure of Part 1 (1924)

5.4 A Scientist's History

Scientists rarely investigate carefully the often only small and gradual conceptual transformations that their insights undergo in the course of historical development, often at the hands of others. Instead of undertaking such a demanding enterprise with little promise of new scientific results, they rather tend to hold onto their insights, reinterpreting them in the light of their present and prospective uses rather than in the light of past achievements, let alone failures. As we shall see, this tendency was inescapable for Hilbert, who understood the progress of physics in terms of an elaboration of the apparently universal and immutable concepts of classical physics.

Indeed, Hilbert described the 1924 Part 1 version of his theory not as a revision of his 1916 Paper 1 version, including major conceptual adjustments and a reorganization of its deductive structure, but essentially as a reprint of his earlier work:

> What follows is essentially a reprint of both of my earlier communications on the *Grundlagen der Physik*, and my comments on them, which were published by F. Klein in his communication *Zu Hilberts erster Note über die Grundlagen der Physik*, with only minor editorial differences and transpositions in order to facilitate their understanding.[153]

Indeed, the organization of Part 1 has not undergone major changes as compared to Paper 1, but seems to represent simply a tightening up; it can be subdivided into the following sections:

1. General Introduction (Hilbert 1924, 1–2)
2. Basic Setting (Hilbert 1924, 2–4)

   Axioms I and II, field equations of electromagnetism and gravitation
3. Basic Theorems (Hilbert 1924, 4–7)

   Theorems 1 (previously II) and 2 (previously III), the theorem earlier designated as Theorem I (now without numbering)
4. Implications for Electromagnetism, Gravitational Field Equations, and Energy-momentum Conservation (Hilbert 1924, 7–11)

   The character of the gravitational part of the Lagrangian, Axiom III (the split of the Lagrangian and the character of the electrodynamical part of the Lagrangian), the gravitational field equations, the form of Mie's Lagrangian, the relation between Mie's energy tensor and Mie's Lagrangian, energy-momentum conservation, and the relation between electromagnetic and gravitational field equations.

The most noteworthy changes in the order of presentation are: a new introductory section and the integration of the treatment of energy-momentum conservation with other results of Hilbert's theory towards the end. Another conspicuous change is that

153 "Das Nachfolgende ist im wesentlichen ein Abdruck der beiden älteren Mitteilungen von mir über die *Grundlagen der Physik* und meiner Bemerkungen dazu, die F. Klein in seiner Mitteilung *Zu Hilberts erster Note über die Grundlagen der Physik* veröffentlicht hat—mit nur geringfügigen redaktionellen Abweichungen und Umstellungen, die das Verständnis erleichtern sollen." (Hilbert 1924, 1)

Hilbert's *Leitmotiv*, Theorem I of Paper 1, has now lost its central place despite meanwhile having been proven by Emmy Noether. As we have seen, even in Paper 1 it no longer played the key heuristic role for Hilbert that it had originally in the Proofs. As the preceding discussion made clear, the rather unchanged form of its presentation hides major changes in the substance of his theory.

These changes are reflected in the introductory section, in a way that again downplays them.

While earlier Hilbert had introduced his own contribution as a solution to the problems raised by Mie and Einstein (Proofs) or Einstein and Mie (Paper 1), he now characterized his results as providing a simple and natural representation of Einstein's general theory of relativity, completed in formal aspects:

> The vast complex of problems and conceptual structures of Einstein's general theory of relativity now find, as I explained in my first communication, their simplest and most natural expression and, in its formal aspect, a systematic supplementation and completion by following the route trodden by Mie.[154]

In view of the overwhelming contemporary impact of Einstein's theory, Mie's role was downplayed in Hilbert's new version. Mie is no longer portrayed as posing problems of a similar profundity to those of Einstein, but as inspiring Hilbert's "simplest and most natural" presentation of general relativity, as well as "a systematic supplementation and completion in its formal aspect."

Instead of attributing a specific role in contemporary scientific discussions to Mie, Hilbert elevates him to the role of one of the founding fathers of a unified-field theoretical worldview:

> The mechanistic ideal of unity in physics, as created by the great researchers of the previous generation and still adhered to during the reign of classical electrodynamics, now must be definitively abandoned. Through the creation and development of the field concept, a new possibility for the comprehension of the physical world has gradually taken shape. Mie was the first to show a way that makes accessible to general mathematical treatment this newly risen 'field theoretical ideal of unity' as I would like to call it.[155]

Curiously neither Einstein nor Minkowski are mentioned in Hilbert's discussion of the spacetime continuum as the "foundation" of "the new field-theoretical ideal":

---

154 "Die gewaltigen Problemstellungen und Gedankenbildungen der allgemeinen Relativitätstheorie von Einstein finden nun, wie ich in meiner ersten Mitteilung ausgeführt habe, auf dem von Mie betretenen Wege ihren einfachsten und natürlichsten Ausdruck und zugleich in formaler Hinsicht eine systematische Ergänzung und Abrundung." (Hilbert 1924, 1–2) The changes in Hilbert's theory were accompanied by a change in his attitude to Einstein's achievement, by which he was increasingly impressed: see (Corry 1999a, 522–525).

155 "Das mechanistische Einheitsideal in der Physik, wie es von den großen Forschern der vorangegangenen Generation geschaffen und noch während der Herrschaft der klassischen Elektrodynamik festgehalten worden war, muß heute endgültig aufgegeben werden. Durch die Aufstellung und Entwickelung des Feldbegriffes bildete sich allmählich eine neue Möglichkeit für die Auffassung der physkalischen Welt aus. Mie zeigte als der erste einen Weg, auf dem dieses neuenstandene "feldtheoretische Einheitsideal", wie ich es nennen möchte, der allgemeinen mathematischen Behandlung zugänglich gemacht werden kann." (Hilbert 1924, 1)

> While the old mechanistic conception takes matter itself as a direct starting point and assumes it to be determined by a finite range of discrete parameters; a physical continuum, the so-called spacetime manifold, rather serves as the foundation of the new field-theoretical ideal. While previously universal laws took the form of [ordinary] differential equations with one independent variable, now partial differential equations are their necessary form of expression.[156]

Mie was exalted to the otherwise rather empty heaven of the founding fathers, leaving room for Hilbert's attempts at a unified theory of gravitation and electromagnetism. He generously mentioned other contemporary efforts as off-springs of his own contribution, a view hardly shared by his contemporaries (see below):

> Since the publication of my first communication, significant papers on this subject have appeared: I mention only Weyl's magnificent and profound investigations, and Einstein's communications, filled with ever new approaches and ideas. In the meantime, even Weyl took a turn in his development that led him too to arrive at just the equations I formulated; and on the other hand Einstein also, although starting repeatedly from divergent approaches, differing among themselves, ultimately returns, in his latest publication, to precisely the equations of my theory.[157]

This passage from Hilbert leaves unspecified to which of his equations he is referring. Given his references to Weyl and Einstein, he must mean the two sets of field equations (51) and (52), which are rather obvious ingredients of any attempted unification of gravitation and electromagnetism. The unique feature of his approach, the specific connection he introduced between these two sets of equations (see (83)) constituting the mathematical expression of electrodynamics as a phenomenon following from gravitation, had become highly problematic and was not adopted by either Weyl or Einstein.

Indeed, it was already problematic whether Weyl's and Einstein's attempts at unification were any more fortunate than Hilbert's. In his concluding paragraph, Hilbert himself expressed his doubts, which were based on the rapid progress of quantum physics, on the one hand, and the lack of any concrete physical results of such theories, on the other:

> Whether the pure field theoretical ideal of unity is indeed definitive, and what possible supplements and modifications of it are necessary to enable in particular the theoretical foundation for the existence of negative and positive electrons, as well as the consistent

156 "Während die alte mechanistische Auffassung unmittelbar die Materie selbst als Ausgang nimmt und diese durch eine endliche Auswahl diskreter Parameter bestimmt ansetzt, dient vielmehr dem neuen feldtheoretischen Ideal das physikalische Kontinuum, die sogenannte Raum-Zeit-Mannigfaltigkeit, als Fundament. Waren früher Differenzialgleichungen mit einer unabhängigen Variablen die Form der Weltgesetze, so sind jetzt notwendig partielle Differenzialgleichungen ihre Ausdrucksform." (Hilbert 1924, 1)

157 "Seit der Veröffentlichung meiner ersten Mitteilung sind bedeutsame Abhandlungen über diesen Gegenstand erschienen: ich erwähne nur die glänzenden und tiefsinnigen Untersuchungen von Weyl und die an immer neuen Ansätzen und Gedanken reichen Mitteilungen von Einstein. Indes sowohl Weyl gibt späterhin seinem Entwicklungsgange eine solche Wendung, daß er auf die von mir aufgestellten Gleichungen ebenfalls gelangt, und andererseits auch Einstein, obwohl wiederholt von abweichenden und unter sich verschiedenen Ansätzen ausgehend, kehrt schließlich in seinen letzten Publikationen geradewegs zu den Gleichungen meiner Theorie zurück." (Hilbert 1924, 2)

> development of the laws holding in the interior of the atom—to answer this is the task for the future.[158]

In spite of his doubts, Hilbert was convinced that "his theory" would endure, (see the preceding paragraph), expressing the belief that it was of programmatic significance for future developments. Even if not, at least philosophical benefit could be drawn from it:

> I am convinced that the theory I have developed here contains an enduring core and creates a framework within which there is sufficient scope for the future development of physics in the sense of a field theoretical ideal of unity. In any case, it is also of epistemological interest to see how the few, simple assumptions I put forth in Axioms I, II, III, and IV suffice for the construction of the entire theory.[159]

The fact that his theory is not based exclusively on these axioms, but also depends rather crucially on other physical concepts, such as energy, and that his theory might change in content as well structure if these concepts changes their meaning,—all of this evidently remained outside of Hilbert's epistemological scope.

6. HILBERT'S ADOPTION OF EINSTEIN'S PROGRAM: THE SECOND PAPER AND ITS REVISIONS

6.1 From Paper 1 to Paper 2

When Hilbert published his Paper 1 in early 1916, he still hoped that his unification of electromagnetism and gravitation would provide the basis for solving the riddles of microphysics. He opened his paper announcing:

> I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.[160]

and concluding:

> ... I am also convinced that through the basic equations established here the most intimate, presently hidden processes in the interior of the atom will receive an explanation,

---

158 "Ob freilich das reine feldtheoretische Einheitsideal ein definitives ist, evtl. welche Ergänzungen und Modifikationen desselben nötig sind, um insbesondere die theoretische Begründung für die Existenz des negativen und des positiven Elektrons, sowie den widerspruchsfreien Aufbau der im Atominneren geltenden Gesetze zu ermöglichen,—dies zu beantworten, ist die Aufgabe der Zukunft." (Hilbert 1924, 2)

159 "Ich glaube sicher, daß die hier von mir entwickelte Theorie einen bleibenden Kern enthält und einen Rahmen schafft, innerhalb dessen für den künftigen Aufbau der Physik im Sinne eines feldtheoretischen Einheitsideals genügender Spielraum da ist. Auch ist es auf jeden Fall von erkenntnistheoretischem Interesse, zu sehen, wie die wenigen einfachen in den Axiomen I, II, III, IV von mir ausgesprochenen Annahmen zum Aufbau der ganzen Theorie genügend sind." (Hilbert 1924, 2)

160 "Die genauere Ausführung sowie vor Allem die spezielle Anwendung meiner Grundgleichungen auf die fundamentalen Fragen der Elektrizitätslehre behalte ich späteren Mitteilungen vor." (Hilbert 1916, 395)

> and in particular that generally a reduction of all physical constants to mathematical constants must be possible ...[161]

Clearly, he intended to dedicate a second communication to the physical consequences of his theory. By March 1916 he had submitted a second installment, which was then withdrawn, no trace remaining.[162] What does remain are the notes of Hilbert's SS 1916 and WS 1916/17 Lectures, and his related Causality Lecture. The WS 1916/17 Lectures offer hints of how his theory would lead to a modification of Maxwell's equations near the sources. While this part is clearly still related to Hilbert's original project, the bulk of these notes testify to his careful study of current work by Einstein and others on general relativity, as well as containing original contributions to that project. In the second communication to the Göttingen Academy submitted at the end of December 1916 (hereafter referred to as "Paper 2"), work on general relativity occupied the entire paper (Hilbert 1917). Hilbert's lecture notes are important for understanding the transition from his original aims to Paper 2, as well as the contents of this paper.[163] One of the most remarkable features of these notes is the openness and informality with which Hilbert shares unsolved problems with his students, later explicitly stating that this was a central goal of his lectures:

> In lectures, and above all in seminars, my guiding principle was not to present material in a standard and as smooth as possible way, just to help the students to maintain ordered notebooks. Above all, I tried to illuminate the problems and difficulties and offer a bridge leading to currently open questions. It often happened that in the course of a semester the program of an advanced lecture was completely changed because I wanted to discuss issues in which I was currently involved as a researcher and which had not yet by any means attained their definite formulation.[164]

6.2 The Causality Quandary

The lecture notes make it clear that Hilbert was still in a quandary over the treatment of causality because his Proofs argument against general covariance seemed to remain valid. The bulk of the typescript notes of his SS 1916 Lectures deal with special relativity (which he calls "die kleine Relativität"): kinematics, and vector and

---

161 "... ich bin auch der Überzeugung, daß durch die hier aufgestellten Grundgleichungen die intimsten, bisher verborgenen Vorgänge innerhalb des Atoms Aufklärung erhalten werden und insbesondere allgemein eine Zurückführung aller physikalischen Konstanten auf mathematische Konstanten möglich sein muß ..." (Hilbert 1916, 407)

162 See the discussion in (Sauer 1999, 560 n. 129).

163 The importance of Hilbert's lectures has been emphasized by Leo Corry. See (Corry 2004).

164 "Es war mein Grundsatz, in den Vorlesungen und erst recht in den Seminaren nicht einen eingefahrenen und so glatt wie möglich polierten Wissensstoff, der den Studenten das Führen sauberer Kollegheften erleichtert, vorzutragen. Ich habe vielmehr immer versucht, die Probleme und Schwierigkeiten zu beleuchten und die Brücke zu den aktuellen Fragen zu schlagen. Nicht selten kam es vor, daß im Verlauf eines Semesters das stoffliche Programm einer höheren Vorlesung wesentlich abgeändert wurde, weil ich Dinge behandeln wollte, die mich gerade als Forscher beschäftigten und die noch keineswegs eine endgültige Gestalt gewonnen hatten." (Reidemeister 1971, 79) Translation by Leo Corry.

tensor analysis (pp. 1–66); dynamics (pp. 66–70 and 76–82); and Maxwell's electrodynamics (pp. 70–76 and 84–89). Hilbert then discusses Mie's theory in its original, special-relativistic form (pp. 90–102), and the need to combine it with "Einstein's concept of the general relativity of events" ("des Einstein'schen Gedankens von der allgemeinen Relativität des Geschehens," p. 103). After introducing the metric tensor, he develops the field equations for gravitation and electromagnetism (pp. 103–111). Discussing these equations, he notes that the causality problem remains unsolved:

> These are 14 equations for the 14 unknown functions $g^{\mu\nu}$ and q_h $(\mu, \nu = 1 \ldots 4)$. The causality principle may or may not be satisfied (the theory has not yet clarified this point). In any event, unlike the case of Mie's theory, the validity of this principle cannot be inferred from simple considerations. Of these 14 equations, 4 (e.g., the 4 Maxwell equations) are a consequence of the remaining 10 (e.g., the gravitational equations). Indeed, the remarkable theorem holds that the number of equations following from Hamilton's principle always corresponds to the number of unknown functions, except in the case occurring here, that the integral is an ["a general" added by hand] invariant.[165]

He still had not resolved the causality problem when he continued the lectures during the winter semester. Among other things, the WS 1916/17 Lecture notes contain much raw material for Paper 2. For example, the discussion of causal relations between events in a given spacetime very much resembles the treatment in that paper.[166] Yet the notes do not discuss the causality question for the field equations.

The same answer to this problem presented in Paper 2 is given in the typescript (unfortunately undated) of his Causality Lecture. From its contents, it is reasonable to conjecture that this is Hilbert's first exposition of his newly-found solution. After discussing the problem for his generally-covariant system of equations and constructing an example to illustrate its nature (pp. 1–5), he comments:

> Einstein's old theory now amounts to the addition of 4 non-invariant equations. But this too is mathematically incorrect. Causality cannot be saved in this way.[167]

---

165 "Dies sind 14 Gleichungen für die 14 unbekannten Funktionen $g^{\mu\nu}$ und q_h $(\mu, \nu, h = 1 \ldots 4)$. Das Kausalitätsprinzip kann erfüllt sein, oder nicht (Die Theorie hat diesen Punkt noch nicht aufgeklärt). Jedenfalls lässt sich auf die Gültigkeit dieses Prinzips nicht wie im Falle der Mie'schen Theorie durch einfache Ueberlegungen schliessen. Von diesen 14 Gleichungen sind nämlich 4 (z.B. die 4 Maxwellschen) eine Folge der 10 übrigen (z.B. der Gravitationsgleichungen). Es gilt nämlich der merkwürdige Satz, dass die Zahl der aus dem Hamiltonschen Prinzip fliessenden Gleichungen immer mit der Zahl der unbekannten Funktionen übereinstimmt, ausser in dem hier eintretenden Fall, das unter dem Integral ["eine allgemeine" added by hand] Invariante steht." (SS 1916 Lectures, 110)

166 See Chapter XIII of the notes, *Einiges über das Kausalitätsprinzip in der Physik,* 97–103, and pp. 57–59 of Paper 2, both of which are discussed below.

167 "Die alte Theorie von Einstein läuft nun darauf hinaus, 4 nicht invariante Gleichungen hinzuzufügen. Aber auch dies ist mathematisch falsch. Auf diesem Wege kann die Kausalität nicht gerettet werden" (p. 5). As discussed above, in his *Entwurf* theory Einstein did not first set up a system of generally-covariant equations and then supplement them by non-invariant conditions; but started from non-generally-covariant field equations. But he had considered the possibility described by Hilbert that these equations have a generally-covariant counterpart, from which they could be obtained by imposing non-invariant conditions.

A similar comment appears in Paper 2:

> In his original theory, now abandoned, A. Einstein (*Sitzungsberichte der Akad. zu Berlin*, 1914, p. 1067) had indeed postulated certain 4 non-invariant equations for the $g_{\mu\nu}$, in order to save the causality principle in its old form.[168]

Neither here nor in any later publication does Hilbert repeat the claim in the lecture notes that this procedure (which he himself had followed in the Proofs) is "mathematisch falsch," which strongly suggests that the notes precede Paper 2.

This suggested temporal sequence is confirmed by another pair of passages: In his lecture, Hilbert compares the problem created by general covariance of a system of partial differential equations and that created by parameter invariance in the calculus of variations:

> The difficulty of having to distinguish between a meaningful and a meaningless assertion is also encountered in Weierstrass's calculus of variations. There the curve to be varied is assumed to be given in parametric form, and one then obtains a differential equation for two unknown functions. One then considers only those assertions that remain invariant when the parameter p is replaced by an arbitrary function of p.[169]

This comparison may well have played a significant role in his solution of the causality problem. The corresponding passage in Paper 2 generalizes this comparison:

> In the theory of curves and surfaces, where a statement in a chosen parametrization of the curve or surface has no geometrical meaning for the curve or surface itself, if this statement does not remain invariant under an arbitrary transformation of the parameters or cannot be brought to invariant form; so also in physics we must characterize a statement that does not remain invariant under any arbitrary transformation of the coordinate system as *physically meaningless*.[170]

This argument is so much more general that it is hard to believe that, once he had hit upon it, Hilbert would have reverted to its restricted application to extremalization of curves. So we shall assume the priority of the Causality Lecture notes.

In these notes, Hilbert asserts that the causality quandary can be resolved by an appropriate understanding of physically meaningful statements:

---

168 "In seiner ursprünglichen, nunmehr verlassenen Theorie hatte A. Einstein (Sitzungsberichte der Akad. zu Berlin. 1914 S. 1067) in der Tat, um das Kausalitätsprinzip in der alten Fassung zu retten, gewisse 4 nicht invariante Gleichungen für die $g_{\mu\nu}$ besonders postuliert." (Hilbert 1917, 61)

169 "Auf die Schwierigkeit, zwischen einer sinnvollen und einer sinnlosen Behauptung unterscheiden zu müssen, stösst man übrigens auch in der Weierstrass'schen Variationsrechnung. Dort wird die zu variierende Kurve als in Parametergestalt gegeben angenommen, und man erhält dann eine Differentialgleichung für zwei unbekannte Funktionen. Man betrachtet dann nur solche Aussagen, die invariant bleiben, wenn man den Parameter p durch eine willkürliche Funktion von p ersetzt." (Causality Lecture, 8)

170 "Gerade so wie in der Kurven- und Flächentheorie eine Aussage, für die die Parameterdarstellung der Kurve oder Fläche gewählt ist, für die Kurve oder Fläche selbst keinen geometrischen Sinn hat, wenn nicht die Aussage gegenüber einer beliebigen Transformation der Parameter invariant bleibt oder sich in eine invariante Form bringen läßt, so müssen wir auch in der Physik eine Aussage, die nicht gegenüber jeder beliebigen Transformation des Koordinatensystems invariant bleibt, als *physikalisch sinnlos* bezeichnen." (Hilbert 1917, 61)

> We obtain the explanation of this paradox by attempting to more rigorously grasp the concept of relativity. It does not suffice to say that the laws of the world are independent of the frame of reference, but rather every single assertion about an event or a concurrence of events only then takes on a physical meaning if it is independent of its designation, i.e. when it is invariant.[171]

In the last clause, one hears distant echoes of Einstein's assertion in his expository paper *Die Grundlage der allgemeinen Relativitätstheorie*:

> We allot to the universe four spacetime variables x_1, x_2, x_3, x_4 in such a way that for every point-event there is a corresponding system of values of the variables $x_1 \ldots x_4$. To two coincident point-events there corresponds one system of values of the variables $x_1 \ldots x_4$, i.e. coincidence is characterized by the identity of the co-ordinates. ... As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of coordinates to others, that is to say, we arrive at the requirement of general co-variance.[172]

Perusal of this paper, published on 11 May 1916 and cited in Hilbert's WS 1916/17 Lectures,[173] may well have contributed to his new understanding of the causality problem.

However, Hilbert's interpretation of a physically meaningful statement actually differs from that of Einstein. Einstein had turned the uniqueness problem for solutions of generally-covariant field equations into an argument against the physical significance of coordinate systems. Hilbert attempted to turn the problem into its own solution by *defining* physically meaningful statements as those for which no such ambiguities arise, whether such statements employ coordinate systems or not. In his Causality Lecture, Hilbert claims to demonstrate the validity of the "causality principle," formulated in terms of physically meaningful statements:

> We would like to prove that the causality principle formulated as follows: "All meaningful assertions are a necessary consequence of the preceding ones [see the citation above]" is valid. Only this theorem is logically necessary and, for physics, also completely sufficient.[174]

To establish this principle, he considers an arbitrary set of generally-covariant field equations (which he calls "ein System invarianter Gleichungen") involving the

---

171 "Die Aufklärung dieses Paradoxons erhalten wir, wenn wir nun den Begriff der Relativität schärfer zu erfassen suchen. Man muss nämlich nicht nur sagen, dass die Weltgesetze vom Bezugssystem unabhängig sind, es hat vielmehr jede einzelne Behauptung über eine Begebenheit oder ein Zusammentreffen von Begebenheiten physikalisch nur dann einen Sinn, wenn sie von der Benennung unabhängig, d.h. wenn sie invariant ist." (Causality Lecture, 5–6)

172 "Man ordnet der Welt vier zeiträumliche Variable x_1, x_2, x_3, x_4 zu, derart, dass jedem Punktereignis ein Wertsystem der Variablen $x_1 \ldots x_4$ entspricht. Zwei koinzidierenden Punktereignissen entspricht dasselbe Wertsystem der Variablen $x_1 \ldots x_4$; d. h. die Koinzidenz ist durch die Übereinstimmung der Koordinaten charakterisiert. .... Da sich alle unsere physikalischen Erfahrungen letzten Endes auf solche Koinzidenzen zurückführen lassen, ist zunächst kein Grund vorhanden, gewisse Koordinatensysteme vor anderen zu bevorzugen, d.h. wir gelangen zu der Forderung der allgemeinen Kovarianz." (Einstein 1916a, 776–777)

173 See (WS 1916/17 Lectures, 112).

metric tensor, the electromagnetic potentials, and their derivatives.[175] He specifies the values of these fields and their derivatives on the space-like hypersurface $t = 0$, which he calls "the present" ("die Gegenwart"); and considers coordinate transformations that do not change the coordinates on this hypersurface, but are otherwise arbitrary (except for continuity and differentiability) off the hypersurface ("die Transformation soll die Gegenwart ungeändert lassen"). He then defines a physically meaningful statement as one that is uniquely determined by Cauchy data, intending to thus establish, at the same time, his principle of causality in terms of what one might call "a mathematical response" to the problem of uniqueness in a generally-covariant field theory:

> Only such a [meaningful assertion] is unequivocally determined by the initial values of $g_{\mu\nu}$, q_μ and their derivatives, and in fact these initial values are to be understood as Cauchy boundary-value conditions. It must be accepted that one can prescribe these boundary values arbitrarily, or that one can proceed to a place in the world at the moment in time when the state characterized by these values prevails. The observer of nature is also considered as standing outside these physical laws; otherwise one would arrive at the antinomies of free will.[176]

As this passage makes clear, Hilbert's proposed definition of physically meaningful statements and clarification of the problem of causality is flawed by the still-unrecognized intricacies of the Cauchy problem in general relativity. He evidently failed to realize that the classical notion of freely-choosable initial values no longer works for generally-covariant field equations since some of them function as constraints on the data that can be given on an initial hypersurface, rather than as evolution equations for that data off this surface. The next section discusses Hilbert's treatment of the problem of causality in Paper 2, including further evidence of his failure to fully grasp Einstein's insight that, in general relativity, coordinate systems have no physical significance of their own.

---

174 "Wir wollen beweisen, dass das so formulierte Kausalitätsprinzip: "Alle sinnvollen Behauptungen sind eine notwendige Folge der vorangegangenen [see the citation above]" gültig ist. Dieser Satz allein ist logisch notwendig und er ist auch für die Physik vollkommen ausreichend." (Causality Lecture, 5–6)

175 The original typescript had specified first and second derivatives of the metric and first derivatives of the electromagnetic potentials, but by hand Hilbert added "beliebig hohen" in the first case and deleted "ersten" in the second.

176 "Nur eine solche [sinnvolle Behauptung] ist durch die Anfangswerte der $g_{\mu\nu}$, q_μ und ihrer Ableitungen eindeutig festgelegt und zwar sind diese Anfangswerte als Cauchy'sche Randbedingungen zu verstehen. Dass man diese Randwerte beliebig vorgeben kann, oder dass man sich an eine Stelle der Welt hinbegeben kann, wo der durch diese Werte charakterisierte Zustand in diesem Zeitmoment herrscht, muss hingenommen werden. Der die Natur beobachtende Mensch wird eben als ausserhalb dieser physikalischen Gesetze stehend betrachtet; sonst käme man zu den Antinomien der Willensfreiheit." (Causality Lecture, 6–7)

6.3 Hilbert at Work on General Relativity

Paper 2 shows that Hilbert's original goal of developing a unified gravito-electromagnetic theory, with the aim of explaining the structure of the electron and the Bohr atom, has been modified in the light of the successes of Einstein's purely gravitational program. Hilbert's shift of emphasis in Paper 1 to the primacy of the gravitational field equations must have facilitated his shift to the consideration of the "empty-space" field equations. From Hilbert's perspective, they are just that subclass of solutions to his fourteen "unified" field equations, for which the electromagnetic potentials vanish. This makes them formally equivalent to the sub-class of solutions to Einstein's field equations with a stress-energy tensor that either vanishes everywhere, or at least outside of some finite world-tube containing the sources of the field. This formal equivalence no doubt contributed to the ease with which contemporary mathematicians and physicists assimilated Hilbert's program to Einstein's, treating Paper 2 as a contribution to the development of the general theory of relativity. This is how Hilbert's contribution came to be assimilated to the relativistic tradition, as we shall discuss in more detail below.

Let us now take a look at the six major topics Hilbert treated in Paper 2:

1. measurement of the components of the metric tensor (Hilbert 1917, 53–55);
2. characteristics and bicharacteristics of the Hamilton-Jacobi equation corresponding to the metric tensor (Hilbert 1917, 56–57);
3. causal relation between events in a spacetime with given metric (Hilbert 1917, 57–59);
4. the causality problem for the field equations determining the metric tensor (Hilbert 1917, 59–63);
5. Euclidean geometry as a solution to the field equations—in particular, the investigation of conditions that characterize it as a unique solution (Hilbert 1917, 63–66 and 70); and
6. the Schwarzschild solution, its derivation (Hilbert 1917, 67–70), and determination of the paths of (freely-falling) particles and light rays in it (Hilbert 1917, 70–76).

1) The metric tensor and its measurement: First of all, Hilbert dropped his previous use of one imaginary coordinate, perhaps influenced by Einstein's use of real coordinates, and emphasized that the $g_{\mu\nu}$, now all real, provide the "Massbestimmung einer Pseudogeometrie" (Hilbert 1917, 54). He classified the elements ("Stücke") of all curves: time-like elements measure proper time; space-like elements measure length; and null elements are segments of a light path. He introduced two ideal measuring instruments: a measuring tape ("Maßfaden") for lengths, and a light clock ("Lichtuhr") for proper times. He makes a comment that suggests, in spite of his remarks in Paper 1 and the Causality Lecture (see above), a lingering

belief in some objective significance to the choice of a coordinate system, independently of the metric tensor:

> First we show that each of the two instruments suffices to compute with its aid the values of the $g_{\mu\nu}$ as functions of x_s, just as soon as a definite spacetime coordinate system x_s has been introduced.[177]

He ends with some comments on a possible axiomatic construction ("Aufbau") of the pseudogeometry, suggesting the need for two axioms:

> first an axiom should be established, from which it follows that length resp. proper time must be integrals whose integrand is only a function of the x_s and their first derivatives with respect to the parameter [p, where $x_s = x_s(p)$ is the parametric representation of a curve]; ...
>
> Secondly an axiom is needed whereby the theorems of the pseudo-Euclidean geometry, that is the old principle of relativity, shall be valid in infinitesimal regions;[178]

2) Characteristics and bicharacteristics: Hilbert defined the null cone at each point, and pointed out that the Monge differential equation (Hilbert 1917, 56):

$$g_{\mu\nu}\frac{dx_\mu}{dp}\frac{dx_\nu}{dp} = 0, \tag{129}$$

and the corresponding Hamilton-Jacobi partial differential equation:

$$g^{\mu\nu}\frac{\partial f}{\partial x_\mu}\frac{\partial f}{\partial x_\nu} = 0, \tag{130}$$

determine the resulting null cone field, the geodesic null lines being the characteristics of the first and the bicharacteristics of the second of these equations. The null geodesics emanating from any world point form the null conoid ("Zeitscheide;" many current texts apply the term "null cone" to non flat spacetimes, but we prefer the term "conoid") emanating from that point. He points out that the equation for these conoids are integral surfaces of the Hamilton-Jacobi equation; and that all timelike world lines emanating from a world point lie inside its conoid, which forms their boundary.

These topics, rather briefly discussed in Paper 2, are treated much more extensively in Hilbert's WS 1916/17 Lectures. In many ways Hilbert's discussion in

---

177 "Zunächst zeigen wir, daß jedes der beiden Instrumente ausreicht, um mit seiner Hülfe die Werte der $g_{\mu\nu}$ als Funktion von x_s zu berechnen, sobald nur ein bestimmtes Raum-Zeit-Koordinatensystem x_s eingeführt worden ist." (Hilbert 1917, 55)

178 "erstens ist ein Axiom aufzustellen, auf Grund dessen folgt, daß Länge bez. Eigenzeit Integrale sein müssen, deren Integrand lediglich eine Funktion der x_s und ihrer ersten Ableitungen nach dem Parameter ist; ...
Zweitens ist ein Axiom erforderlich, wonach die Sätze der pseudo-Euklidischen Geometrie d.h. das alte Relativitätsprinzip im Unendlichkleinen gelten soll;" (Hilbert 1917, 56)

Paper 2 reads like a précis of these notes; it becomes much more intelligible if they are consulted. Chapter IX (pp. 69–80) entitled "Die Monge'sche Differentialgleichung" also treats the Hamilton-Jacobi equation and the theory of characteristics, emphasizing their relation to the Cauchy problem, and the reciprocal relation between integral surfaces of the Hamilton-Jacobi equation (the null conoids are called "transzendentale Kegelfläche") and null curves. Chapters X (pp. 80–82, "Die vierdimensionale eigentliche u. Pseudogeometrie") and XI (pp. 82–97, "Zusammenhang mit der Wirklichkeit") cover the material in the first section of Paper 2: the measuring tape ("Massfaden") is discussed in section 38 (pp. 85–86 and pp. 91–92), and the light clock, already introduced in the context of special relativity (see the SS 1916 Lectures, 6–10), is reintroduced in section 44 (pp. 93–94, "Axiomatische Definition der Lichtuhr"). Both instruments are used to determine the components of the metric tensor as functions of the coordinates, "sobald nur ein bestimmtes Raum-Zeit Koordinatensystem x_i eingeführt worden ist" (p. 95).

3) Causal relation between events: [179] In accord with the implicit requirement that three of the coordinates be space-like and one time-like, Hilbert imposes corresponding conditions on the components of the metric tensor. But he has a unique way of motivating them:

> Up to now all coordinate systems x_s that result from any one by arbitrary transformation have been regarded as equally valid. This arbitrariness must be restricted when we want to realize the concept that two world points on the same time line can be related as cause and effect, and that it should then no longer be possible to transform such world points to be simultaneous. In declaring x_4 as the *true* time coordinate we adopt the following definition:
>
> ...
>
> So we see that the concepts of cause and effect, which underlie the principle of causality, also do not lead to any inner contradictions whatever in the new physics, if we only take the inequalities (31) always to be part of our basic equations, that is if we confine ourselves to using *true* spacetime coordinates.[180]

Again, he seems to believe that there is some residual physical significance in the choice of a coordinate system: it must reflect the relations of cause and effect between events on the same time-like world line. He defines a proper ("eigentliches") coordinate system as one, in which (in effect) the first three coordinates are space-like and the fourth time-like in nature; transformations between such proper coordinate systems are also called proper. Given Hilbert's stated goal of restricting the choice of coordinates to those that reflect the causal order on all time-like world lines, his con-

179 This section also includes material from Hilbert's WS 1916/17 Lectures: Chapter XII, *Einiges über das Kausalitätsprinzip in der Physik,* (pp. 97–104) covers the same ground as, but in no more detail than, the text of Paper 2.

ditions are sufficient but not necessary since they exclude retarded null coordinates, which also preserve this causal order.

4) Causality problem for the field equations: As noted, Hilbert's analysis follows his Causality Lecture. In Paper 2 he writes:

> Concerning the principle of causality, let the physical quantities and their time derivatives be known at the present in some given coordinate system: then a statement will only have physical meaning if it is invariant under all those transformations, for which the coordinates just used for the present remain unchanged; I maintain that statements of this type for the future are all uniquely determined, that is, *the principle of causality holds in this form*:
>
> *From present knowledge of the 14 physical potentials* $g_{\mu\nu}$, q_s *all statements about them for the future follow necessarily and uniquely provided they are physically meaningful*.[181]

A hasty reading might suggest that Hilbert is asserting the independence of all physically meaningful statements from the choice of a coordinate system, and he has often been so interpreted; but this is not what he actually says. His very definition of physically meaningful ("physikalisch Sinn haben") involves the class of coordinate systems that leave the coordinates on the initial hypersurface ("die Gegenwart") unchanged. Secondly, Hilbert uses a Gaussian coordinate system, introduced earlier,[182] in order to prove his assertion about the causality principle.[183] Finally, if his words were so interpreted, they would stand in flagrant contradiction to his earlier statements (cited above)

---

180 "Bisher haben wir alle Koordinatensysteme x_s die aus irgend einem durch eine willkürliche Transformation hervorgehen, als gleichberechtigt angesehen. Diese Willkür muß eingeschränkt werden, sobald wir die Auffassung zur Geltung bringen wollen, daß zwei auf der nämlichen Zeitlinie gelegene Weltpunkte im Verhältnis von Ursache und Wirkung zu einander stehen können und daß es daher nicht möglich sein soll, solche Weltpunkte auf gleichzeitig zu transformieren.
...
So sehen wir, daß die dem Kausalitätsprinzip zu Grunde liegenden Begriffe von Ursache und Wirkung auch in der neuen Physik zu keinerlei inneren Widersprüche führen, sobald wir nur stets die Ungleichungen (31) [the conditions Hilbert imposes on the metric tensor] zu unseren Grundgleichungen hinzunehmen d.h. uns auf den Gebrauch *eigentlicher* Raum-zeitkoordinaten beschränken." (Hilbert 1917, 57 and 58)

181 "Was nun das Kausalitätsprinzip betrifft, so mögen für die Gegenwart in irgend einem gegebenen Koordinatensystem die physikalischen Größen und ihre zeitlichen Ableitungen bekannt sein: dann wird eine Aussage nur physikalisch Sinn haben, wenn sie gegenüber allen denjenigen Transformationen invariant ist, bei denen eben die für die Gegenwart benutzten Koordinaten unverändert bleiben; ich behaupte, daß die Aussagen dieser Art für die Zukunft sämtlich eindeutig bestimmt sind d.h. das Kausalitätsprinzip gilt in dieser Fassung:
Aus der Kenntnis der 14 physikalischen Potentiale $g_{\mu\nu}$, q_s *in der Gegenwart folgen alle Aussagen über dieselben für die Zukunft notwendig und eindeutig, sofern sie physikalischen Sinn haben*." (Hilbert 1917, 61)

182 See (Hilbert 1917, 58–59).

183 See (Hilbert 1917, 61–62).

about the measurement of the metric and the causal relation between events which presuppose attaching some residual physical meaning to the choice of coordinates.

His proof consists of a brief discussion of the Cauchy problem for the field equations in a Gaussian coordinate system. One of us has discussed this aspect of his work elsewhere (Stachel 1992), so we shall be brief here. He only considers the ten gravitational field equations (51) since he interprets Theorem I of Paper 1 as showing that the other four (52) follow from them. Gaussian coordinates eliminate four of the 14 field quantities, the $g_{0\mu}$, leaving only ten (the six g_{ab}, $a, b = 1, 2, 3$, and the four q_s), so he concludes that the resulting system of equations is in Cauchy normal form. This treatment is erroneous on several counts, but we postpone discussion of this question until the next section. More relevant to the present topic is Hilbert's statement:

> Since the Gaussian coordinate system itself is uniquely determined, therefore also all statements about those potentials (34) [the ten potentials mentioned above] with respect to these coordinates are of invariant character.[184]

He never discusses the behavior of the initial data under coordinate transformations on the initial hypersurface (three-dimensional hypersurface diffeomorphisms in modern terminology), confirming that his treatment is still tied to the use of particular coordinate systems rather than being based on coordinate-invariant quantities.

Finally, his discussion of how to implement the requirement of physically meaningful assertions depends heavily on the choice of a coordinate system. He remarks:

> The forms, in which physically meaningful, i.e. invariant, statements can be expressed mathematically are of great variety.[185]

and proceeds to discuss three ways:

> *First*. This can be done by means of an invariant coordinate system. ...
>
> *Second*. The statement, *according to which a coordinate system can be found* in which the 14 potentials $g_{\mu\nu}$, q_s have certain definite values in the future, or fulfill certain definite conditions, is always an invariant and therefore a physically meaningful one. ...
>
> *Third*. A statement is also invariant and thus has physical meaning if it is supposed to be valid in any arbitrary coordinate system.[186]

---

184 "Da das Gaußische Koordinatsystem selbst eindeutig festgelegt ist, so sind auch alle auf dieses Koordinatensystem bezogenen Aussagen über jene Potentiale (34) von invariantem Charakter." (Hilbert 1917, 62)

185 "Die Formen in denen physikalisch sinnvolle d.h. invariante Aussagen mathematisch zum Ausdruck gebracht werden können, sind sehr mannigfaltig." (Hilbert 1917, 62)

186 "Erstens. Dies kann mittelst eines invarianten Koordinatensystem geschehen. ...
Zweitens. Die Aussage, wonach sich ein Koordinatensystem finden läßt, in welchem die 14 Potentiale $g_{\mu n}$, q_s für die Zukunft gewisse bestimmte Werte haben oder gewisse Beziehungen erfüllen, ist stets eine invariante und daher physikalisch sinnvoll. ...
Drittens. Auch ist eine Aussage invariant und hat daher stets physikalisch Sinn, wenn sie für jedes beliebige Koordinatensystem gültig sein soll." (Hilbert 1917, 62–63)

The first two ways explicitly depend on the choice of a coordinate system, which is not necessarily unique. As examples of the first way, he cites Gaussian and Riemannian coordinates. It is true that, discussing the second, he notes:

> The mathematically invariant expression for such a statement is obtained by eliminating the coordinates from those relations.[187]

But he does not give an example, nor does he suggest the most obvious way of realizing his goal, if indeed it was a coordinate-independent solution to the problem: the use of invariants as coordinates. As Kretschmann noted a few years later, in matter- and field-free regions the four non-vanishing invariants of the Riemann tensor may be used as coordinates. If the metric is then expressed as a function of these coordinates, its components themselves become invariants.[188] The use of such coordinates was taken up again by Arthur Komar in the 1960s, and today they are often called Kretschmann-Komar coordinates.[189]

One might think that Hilbert had in mind something like this in his third suggested way. However, the example he cites makes it clear that he meant something else:

> An example of this are Einstein's energy-momentum equations having divergence character. For, although Einstein's energy [that is, the gravitational energy-momentum pseudotensor] does not have the property of invariance, and the differential equations he put down for its components are by no means covariant as a system of equations, nevertheless the assertion contained in them, that they shall be satisfied in any coordinate system, is an invariant demand and therefore it carries physical meaning.[190]

Rather than invariant quantities, evidently he had in mind non-tensorial entities and sets of equations, which nevertheless take the same form in every coordinate system.

In summary, Hilbert's treatment in Paper 2 of the problem of causality in general relativity still suffers from many of the flaws in his original approach. In particular, physical significance is still ascribed to coordinate systems, and the claim is maintained that the identities following from Theorem I represent a coupling between the two sets of field equations. On the other hand, his efforts to explore the solutions of the gravitational field equations from the perspective of a mathematician produced significant contributions to general relativity, to be discussed later.

---

187 "Der mathematische invariante Ausdruck für eine solche Aussage wird durch Elimination der Koordinaten aus jenen Beziehungen erhalten." (Hilbert 1917, 62–63)

188 See (Kretschmann 1917).

189 See (Komar 1958).

190 "Ein Beispiel dafür sind die Einsteinschen Impuls-Energiegleichungen vom Divergenz Character. Obwohl nämlich die Einsteinsche Energie die Invarianteneigenschaft nicht besitzt und die von ihm aufgestellten Differentialgleichungen für ihre Komponenten auch als Gleichungssystem keineswegs kovariant sind, so ist doch die in ihnen enthaltene Aussage, daß sie für jedes beliebige Koordinatensystem erfüllt sein sollen, eine invariante Forderung und hat demnach einen physikalischen Sinn." (Hilbert 1917, 63)

5) Euclidean geometry: This section opens with some extremely interesting general comments contrasting the role of geometry in what Hilbert calls the old and the new physics:

> The old physics with the concept of absolute time took over the theorems of Euclidean geometry and without question put them at the basis of every physical theory. ...
>
> The new physics of Einstein's principle of general relativity takes a totally different position vis-à-vis geometry. It takes neither Euclid's nor any other particular geometry *a priori* as basic, in order to deduce from it the proper laws of physics, but, as I showed in my first communication, the new physics provides at one fell swoop through one and the same Hamilton's principle the geometrical and the physical laws, namely the basic equations (4) and (5) [the ten gravitational and four electromagnetic field equations], which tell us how the metric $g_{\mu\nu}$ —at the same time the mathematical expression of the phenomenon of gravitation—is connected with the values q_s of the electrodynamic potentials.[191]

Hilbert declares:

> With this understanding, an old geometrical question becomes ripe for solution, namely whether and in what sense Euclidean geometry—about which we know from mathematics only that it is a logical structure free from contradictions—also possesses validity in the real world.[192]

He later formulates this question more precisely:

> The geometrical question mentioned above amounts to the investigation, whether and under what conditions the four-dimensional Euclidean pseudo-geometry [i.e., the Minkowski metric] ... is a solution, or even the only regular solution, of the basic physical equations.[193]

Hilbert thus takes up a problem that emerged with the development of non-Euclidean geometry in the 19th century and considered by such eminent mathematicians as Gauss and Riemann: the question of the relation between geometry and physical real-

---

191 "Die alte Physik mit dem absoluten Zeitbegriff übernahm die Sätze der Euklidische Geometrie und legte sie vorweg einer jeden speziellen physikalischen Theorie zugrunde. ...
Die neue Physik des Einsteinschen allgemeinen Relativitätsprinzips nimmt gegenüber der Geometrie eine völlig andere Stellung ein. Sie legt weder die Euklidische noch irgend eine andere bestimmte Geometrie vorweg zu Grunde, um daraus die eigentlichen physikalischen Gesetze zu deduzieren, sondern die neue Theorie der Physik liefert, wie ich in meiner ersten Mitteilung gezeigt habe, mit einem Schlage durch ein und dasselbe Hamiltonsche Prinzip die geometrischen und die physikalischen Gesetze nämlich die Grundgleichungen (4) und (5), welche lehren, wie die Maßbestimmungen $g_{\mu\nu}$ — zugleich der mathematischen Ausdruck der physikalischen Erscheinung der Gravitation — mit den Werten q_s der elektrodynamischen Potentiale verkettet ist." (Hilbert 1917, 63–64)

192 "Mit dieser Erkenntnis wird nun eine alte geometrische Frage zur Lösung reif, die Frage nämlich, ob und in welchem Sinne die Euklidische Geometrie — von der wir aus der Mathematik nur wissen, daß sie ein logisch widerspruchsfreier Bau ist — auch in der Wirklichkeit Gültigkeit besitzt." (Hilbert 1917, 63)

193 "Die oben genannte geometrische Frage läuft darauf hinaus, zu untersuchen, ob und unter welchen Voraussetzungen die vierdimensionale Euklidische Pseudogeometrie ... eine Lösung der physikalischen Grundgleichungen bez. die einzige reguläre Lösung derselben ist." (Hilbert 1917, 64)

ity. For a number of reasons, this question was not central to Einstein's heuristic. He had never addressed the question posed by Hilbert: the conditions under which Minkowski spacetime is a unique solution to the gravitational field equations. To Einstein, the question of the Newtonian limit, and hence the incorporation of Newton's theory into his new theory of gravitation, was much more important than the question of the existence of matter-free solutions to his equations. Indeed, this question was a rather embarrassing one for Einstein since such solutions display inertial properties of test particles even in the absence of matter, a feature that he had difficulty in accepting because of his Machian conviction that all inertial effects must be due to interaction of masses.[194] By establishing a connection between general relativity and the mathematical tradition questioning the geometry of physical space, Hilbert made a significant contribution to the foundations of general relativity.

In attempting to answer the question of the relation between Minkowski spacetime and his equations, Hilbert first of all notes that, if the electrodynamic potentials vanish, then the Minkowski metric is a solution of the resulting equations, i.e., of the vanishing of what we now call the Einstein tensor.[195] He then poses the converse question: under what conditions is the Minkowski metric the *only* regular solution to these equations? He considers small perturbations of the Minkowski metric (a technique that Einstein had already introduced) and shows that, if these perturbations are time independent (curiously, here reverting to use of an imaginary time coordinate) and fall off sufficiently rapidly and regularly at infinity, then they must vanish everywhere. In the next section of the paper, he proves another relevant result, which we shall discuss below.

This section of Paper 2 is a condensation of material covered in his WS 1916/17 Lectures:

- in the table of contents (p. 197), pp. 104–106 are entitled: "Der Sinn der Frage: Gilt die Euklidische Geometrie?"
- pp. 109–111 are headed "Gilt die Euklidische Geometrie in der Physik?" in the typescript, with the handwritten title "Die Grundgleichungen beim Fehlen von Materie" added in the margin, and entitled "Aufstellung der Grundgleichungen beim Fehlen der Materie" in the table of contents; and
- pp. 111–112, bear the handwritten title "Zwei Sätze über die Gültigkeit der Euklidischen Geometrie" in the margin, and "Zwei noch unbewiesene Sätze über die Gültigkeit der Pseudoeuklidischen Geometrie in der Physik" in the table of contents.

The lecture notes make much clearer than Paper 2 Hilbert's motivation for a discussion of the empty-space field equations in general, and of the Schwarzschild metric in particular. In the notes, Hilbert introduces the field equations in section 51 (WS

194 For a historical discussion, see (Renn 1994).

195 "wenn alle Elektrizität entfernt ist, so ist die pseudo-Euklidische Geometrie möglich" See (Hilbert 1917, 64).

1916/17 Lectures, 106–109),[196] sandwiched between discussions of his motivation for raising the question of the validity of Euclidean geometry and his attempts to answer it. At the end of the previous section he points out:

> We would like to anticipate the results of our calculation: in general our basic physical equations have no solutions at all. In my opinion, this is a positive result of the theory: since in no way are we able to impose Euclidean geometry on nature through a different interpretation of experiments. Assuming namely that my basic physical equations to be developed are really correct, then no other physics is possible, i.e., reality cannot be understood in a different way.[197]

Hilbert evidently thought he had found a powerful argument against geometric conventionalism—presumably, he had Poincaré in mind here. He continues:

> On the other hand we shall see that under certain very specialized assumptions—perhaps the absence of matter throughout space is sufficient for this—the only solution to the differential equations are $g_{\mu\nu} = \delta_{\mu\nu}$ [the Minkowski metric].[198]

At this point, the problem of the status of geometry is broadened from three-dimensional geometry to four dimensional pseudo-geometry—and in particular the question of the status of Euclidean geometry is broadened to that of four-dimensional Minkowski pseudo-geometry. In this form, it plays a central role in Hilbert's thinking about his program. This problem, rooted as it was in a mathematical tradition going back to Gauss, led him naturally to consider what we call the empty-space Einstein field equations. He hoped that the absence of matter and non-gravitational fields might suffice to uniquely single out the Minkowski metric as a solution to his field equations (which are identical to Einstein's in this case):

> It is possible that the following theorem is correct:
> Theorem: If one removes all electricity from the world (i.e. $q_i \equiv 0$) and demands absolute regularity—i.e. the possibility of expansion in a power series—of the gravitational potentials $g_{\mu\nu}$ (a requirement that in our opinion must always be fulfilled, even in the general case), then Euclidean geometry prevails in the world, i.e. the 10 equations (3) [equation number in the original; the vanishing of the Einstein tensor] have $g_{\mu\nu} = \delta_{\mu\nu}$ as their only solution.[199]

(He explains what he means by "regular" in his discussion of the Schwarzschild metric, considered below.) Of course, Hilbert was *not* able to establish this theorem, since it is not true, as Einstein's work on gravitational waves might already have sug-

---

196 Page 107 is missing from the typescript.

197 "Wir wollen das Resultat unserer Rechnung vorwegnehmen: unsere physikalischen Grundgleichungen haben im allgemeinen keineswegs Lösungen. Dies ist meiner Meinung nach ein positives Resultat der Theorie: denn wir können der Natur die Euklidischen Geometrie durch andere Deutung der Experimente durchaus nicht aufzwingen. Vorausgesetzt nämlich, dass meine zu entwickelnden physikalischen Grundgleichungen wirklich richtig sind, so ist auch keine andere Physik möglich, d.h., die Wirklichkeit kann nicht anders aufgefasst werden." (WS 1916/17 Lectures, 106)

198 "Andererseits werden wir sehen, dass unter gewissen sehr spezialisierenden Voraussetzungen—vielleicht ist das Fehlen von Materie im ganzen Raum dazu schon hinreichend—die einzige Lösungen der Diffentialgleichungen $g_{\mu\nu} = \delta_{\mu\nu}$ [the Minkowski metric] sind."

gested (Einstein 1916c). Nor was he able to find any other set of necessary and sufficient conditions for the uniqueness of the Minkowski metric; but he did almost establish one set of sufficient conditions and proved another:

> I consider the following theorem to be very probably correct: If one removes all electricity from the world and demands for the gravitational potential, apart from the self-evident requirement of regularity, that $g_{\mu\nu}$ is independent of t, i.e. that gravitation is static, and finally [one demands] also regular behavior at infinity, then $g_{\mu\nu} = \delta_{\mu\nu}$ are the only solutions to the gravitation equations (3)[equation number in the original].
> I can now already prove this much of the theorem, that in the neighborhood of Euclidean geometry there are certainly no solutions to these equations.[200]

This is, of course, the result that he did prove in Paper 2 (see above). The proof of this result for the full, non-linear field equations hung fire for a long time with several proofs for the case of static metrics being given over the years; the proof for stationary metrics was finally given by André Lichnerowicz in 1946.[201]

6) The Schwarzschild solution: The Schwarzschild solution had already been published (Schwarzschild 1916) and Hilbert dedicates considerable space to it, both in his lecture notes and in Paper 2. He uses it in the course of his effort to exploit the new tools of general relativity for addressing the foundational questions of geometry raised in the mathematical tradition. In his lecture notes, he introduces a number of assumptions on the metric tensor in order to prove a theorem on the uniqueness of Euclidean geometry:

> 1) Let $g_{\mu\nu}$ again be independent of t.
>
> 2) Let $(g_{\nu 4} = 0)$ $(\nu = 1, 2, 3)$ [interpolated by hand: "i.e. Gaussian coordinate system, which can always be introduced by a transformation"] (Orthogonality of the t- axis to the x_1, x_2, x_3 -space, the so-called metric space.)
>
> 3) There is a distinguished point in the world, with respect to which central symmetry holds, i.e. the rotation of the coordinate system around this point is a transformation of the world onto itself.

---

199 "Es ist möglich, dass folgender Satz richtig ist:
Satz: Nimmt man alle Elektrizität aus der Welt hinweg (d.h. $q_i \equiv 0$) und verlangt man absolute Regularität—d.h. Möglichkeit der Entwicklung in eine Potenzreihe—der Gravitationspotentiale $g_{\mu\nu}$ (eine Forderung, die nach unserer Auffassung auch im allgemeinen Fall immer erfüllt sein muss), so herrscht in der Welt die Euklidische Geometrie, d.h. die 10 Gleichungen (3) haben $g_{\mu\nu} = \delta_{\mu\nu}$ als einzige Lösung." (WS 1916/17 Lectures, 111–112)

200 "Für sehr wahrscheinlich richtig halte ich folgenden Satz:
Nimmt man alle Elektrizität aus der Welt fort und verlangt von den Gravitationspotentialen ausser der selbstverständlichen Forderung der Regularität noch, dass $g_{\mu\nu}$ von t unabhängig ist, d.h. dass die Gravitation stille steht, und schliesslich noch reguläres Verhalten im Unendlichen, so sind $g_{\mu\nu} = \delta_{\mu\nu}$ die einzigen Lösungen der Gravitationsgleichungen (3).
Von diesem Satz kann ich schon jetzt so viel beweisen, dass in der Nachbarschaft der Euklidischen Geometrie sicher keine Lösung dieser Gleichungen vorhanden sind." (WS 1916/17 Lectures, 112)

201 See (Lichnerowicz 1946).

> Now the following theorem holds: If the gravitational potentials fulfill conditions 1–3, then Euclidean geometry is the only solution to the basic physical equations.[202]

The proof of this theorem leads him to consider the problem of spherically-symmetric solutions to the empty-space Einstein field equations, a problem that Hilbert notes had previously been treated by Einstein (in the linear approximation) and Schwarzschild (exactly). He claims for his own calculations only that, compared to those of others, they are "auf ein Minimum reduziert" (WS 1916/17 Lectures, 113) by working from his variational principle for the field equations (see above). Hermann Weyl gave a similar variational derivation in 1917 (Weyl 1917); the section of his book *Raum-Zeit-Materie* on the Schwarzschild metric includes a reference to Hilbert's Paper 2, which reproduces Hilbert's variational derivation, (Weyl 1918a; 1918b, 230 n.9; 1923, 250 n.19). But Pauli's magisterial survey of the theory of relativity mentions only Weyl's paper, this probably contributing to the neglect of Hilbert's contribution in most later discussions (Pauli 1921).

In Paper 2, Hilbert derives the Schwarzschild metric from the same three assumptions as in the lecture notes, emphasizing that:

> In the following I present for this case a procedure that makes no assumptions about the gravitational potentials $g_{\mu\nu}$ at infinity, and which moreover offers advantages for my later investigations.[203]

In spite of this, many later derivations of the Schwarzschild metric still continue to impose unnecessary boundary conditions. But Hilbert did not show that the assumption of time-independence is also unnecessary, as proved by Birkhoff in 1923. (The assertion that the Schwarzschild solution is the only spherically symmetric solution to the empty-space Einstein equations is known as Birkhoff's theorem.)[204]

Hilbert's discussion of the Schwarzschild solution also raises the problem of its singularities and their relation to Hilbert's theory of matter. In his lecture notes, after establishing the Schwarzschild metric, he writes:

---

202 "1) Es sei wieder $g_{\mu\nu}$ unabhängig von t.
2) Es sei $g_{\nu 4} = 0 \qquad \nu = 1, 2, 3$ [interpolated by hand: "d.h. Gauss'sches Koordinatensystem, das durch Transformation immer eingeführt werden kann"] (Orthogonalität der t-Achse auf dem x_1, x_2, x_3 -Raum, dem sogenannten Streckenraum.)
3) Es gebe einen ausgezeichneten Punkt in der Welt, in Bezug auf welchen zentrische Symmetrie vorhanden sein soll, d.h. die Drehung des Koordinatensystems um diesen Punkt ist eine Transformation der Welt in sich.
Nun gilt folgender Satz:
Erfüllen die Gravitationspotentiale die Bedingungen 1–3, so ist die Euklidische Geometrie die einzige Lösung der physikalischen Grundgleichungen." (WS 1916/17 Lectures, 113)

203 "Ich gebe im Folgenden für diesen Fall einen Weg an, der über die Gravitationspotentiale $g_{\mu\nu}$ im Unendlichen keinerlei Voraussetzungen macht und ausserdem für meine späteren Untersuchungen Vorteile bietet." (Hilbert 1917, 67) For the derivation, see pp. 67–70.

204 See (Birkhoff 1923, 253–256).

> According to our conception of the nature of matter, we can only consider those $g_{\mu\nu}$ to be physically viable solutions to the differential equations $K_{\mu\nu} = 0$ [the Einstein equations] that are regular and singularity free.
> We call a gravitational field or a metric "regular"—this definition had to be added—when it is possible to introduce a coordinate system, such that the functions $g_{\mu\nu}$ are regular and have a non-zero determinant at every point in the world. Furthermore, we describe a single function as being regular if it and all its derivatives are finite and continuous. This is incidentally always the definition of regularity in physics, whereas in mathematics a regular function is required to be analytic.[205]

It is curious that Hilbert identifies physical regularity with *infinite* differentiability and continuity of all derivatives. Either of these requirements is much too strong: each precludes gravitational radiation carrying new information, for example gravitational shock waves.[206] But at least Hilbert attempted to define a singularity of the gravitational field. In his understanding, the Schwarzschild solution has singularities at $r = 0$ and at the Schwarzschild radius. But we now know the first singularity is real, while the second can be removed by a coordinate transformation. He remarks:

> When we consider that these singularities are due to the presence of a mass, then it also seems plausible that they cannot be eliminated by coordinate transformations. However, we will give a rigorous proof of this later by examining the behavior of geodesic lines in the vicinity of this point.[207]

Hilbert then returns to his original motif: the Schwarzschild solution as a tool for discussing foundational problems of geometry:

> In order to obtain singularity-free solutions, we must assume that a [i.e., the mass parameter] = 0. [This leads to the Minkowski metric.] ... This proves the ... theorem: In the absence of matter, under the stated assumptions 1–3 [see above], the pseudo-Euclidean geometry of the little relativity principle [i.e., special relativity] actually holds in physics; and for t = const Euclidean geometry is in fact realized in the world.[208]

---

205 "Nach unserer Auffassung vom Wesen der Materie können wir als physikalisch realisierbare Lösungen $g_{\mu\nu}$ der Differentialgleichungen $K_{\mu\nu} = 0$ [the Einstein equations] nur diejenigen ansehen, welche regulär und singularitätenfrei sind.
"Regulär" nennen wir ein Gravitationsfeld oder eine Massbestimmung,— diese Definition war noch nachzutragen—wenn es möglich ist, ein solches Koordinatensystem einzuführen, dass die Funktionen $g_{\mu\nu}$ an jeder Stelle der Welt regulär sind und eine von null verschiedenen Determinante haben. Wir bezeichnen ferner eine einzelne Funktion als regulär, wenn sie mit allen ihren Ableitungen endlich und stetig ist. Dies ist übrigens immer die Definition der Regularität in der Physik, während in der Mathematik von einer regulären Funktion verlangt wird, dass sie analytisch ist." (WS 1916/17 Lectures, 118)

206 See, e.g., (Papapetrou 1974, 169–177).

207 "Wenn wir bedenken, dass diese Singularitäten von der Anwesenheit einer Masse herrühren, so erscheint es auch plausibel, dass dieselben durch Koordinatentransformation nicht zu beseitigen sind. Einen strengen Beweis dafür werden wir aber erst weiter unten geben, indem wir den Verlauf der geodätischen Linien in der Umgebung dieser Punkt untersuchen." (WS 1916/17 Lectures, 118–119)

208 "Wir müssen also, um singularitätenfreie Lösungen zu erhalten, a [i.e., the mass parameter] = 0 annehmen. Wir haben damit den ... Satz bewiesen: Bei Abwesenheit von Materie ($q_i = 0$) existiert unter den ... genannten Voraussetzungen 1–3 [see above] die pseudoeuklidischen Geometrie des kleinen Relativitätsprinzips in der Physik tatsächlich, und für t = const ist in der Welt die Euklidische Geometrie wirklich realisiert." (WS 1916/17 Lectures, 119)

In the sequel, Hilbert explores its physical significance for describing the behavior of matter in space and time. His conception of matter, based on Mie's theory, plays no significant role in this discussion, its role being taken instead by assumptions that Hilbert assimilated from Einstein's work, such as the geodesic postulate for the motion of free particles.

He then turns to the justification for considering the case $a \neq 0$:

> Then we are acting against our own prescription that we shall regard only singularity-free gravitational fields as realizable in nature. Hence we must justify the assumption $a \neq 0$. [209]

He emphasizes the extraordinary difficulty of integrating the 14 field equations, even for "the simple special case when they go over to $K_{\mu\nu} = 0$":

> Mathematical difficulties already hinder us, for example, from constructing a single neutral mass point. If we were able to construct such a neutral mass, and if its behavior in the neighborhood of this point were known, then, if we let the neutral mass degenerate increasingly to a mass point, the $g_{\mu\nu}$ at this point would display a singularity. Such a singularity we would have to regard as being allowed in the sense that the $g_{\mu\nu}$ outside the immediate neighborhood of the singularity correctly describes the course actually realized in nature. In [the Schwarzschild line element] we must now have this kind of singularity at hand. Incidentally, we can now state that the construction of a neutral mass point, even if this is possible later, will prove to be so complicated that for purposes, in which one does not look at the immediate neighborhood of the mass point, one will be able to calculate the approximately correct gravitational potentials containing a singularity with sufficient precision.
>
> We now maintain the following: If we could actually carry out the mathematical expansion leading to construction of a neutral massive particle, we would probably find laws that, for the time being, still must be formulated axiomatically; but which later will emerge as consequences of our general theory, consequences that admittedly only can be proven categorically by means of a broad-ranging theory and complex calculations. These axioms, which thus have only provisional significance, we formulate as follows:
>
> Axiom I.: The motion of a mass point in the gravitational field is represented by a geodesic line that is a time-like.
>
> Axiom II: The motion of light in the gravitational field is represented by a null geodesic curve.
>
> Axiom III.: A singular point of the metric is equivalent to a gravitational center.[210]

Hilbert calls the first two axioms, taken from Einstein's work, a "rational generalization" of the behavior of massive particles and light rays in the "old physics," in which the metric tensor takes the limiting Minkowski values. He states that the Newtonian law of gravitational attraction and the resulting Keplerian laws of planetary

---

209 "Dann handeln wir zwar entgegen unserer eigenen Vorschrift, dass wir nur singularitätenfreie Gravitationsfelder als in der Natur realisierbar ansehen wollen. Daher müssen wir die Annahme $a \neq 0$ rechtfertigen." (WS 1916/17 Lectures, 120)

motion follow from these axioms "in the first approximation." In this way, Hilbert integrated into his theory the essential physical elements, on which Einstein's path to general relativity was based. Even his epistemological justification for the superiority of the new theory now makes use of an argument for the integration of knowledge. Remarkably, from Hilbert's perspective, this integration not only involves knowledge of classical physics such as Newton's law of gravitation, but also of Euclidean geometry as a physical interpretation of space:

> In principle, however, this new Einsteinian law has no similarity to the Newtonian. It is infinitely more complicated than the latter. If we nevertheless prefer it to the Newtonian, this is because this law satisfies a profound philosophical principle—that of general invariance—and that it contains as special cases two such heterogeneous things as on the one hand, Newton's law and on the other, the actual validity of Euclidian geometry in physics under certain simple conditions; so that we do not have to, as was the case until now, first assume the validity of Euclidian geometry and then put together a law of attraction.[211]

Thus we see that Hilbert considers his results on the conditions of validity of Euclidean geometry on a par in importance with, and logically prior to, Einstein's and Schwarzschild's results on the Newtonian limit of general relativity.

In accord with the physical interpretation they are given in Axioms I and II, Hilbert then goes on to study the time-like and null geodesics of the Schwarzschild metric, leading to discussions of two general-relativistic effects that Einstein had already considered: the planetary perihelion precession and the deflection of light due to the Sun's gravitational field. This discussion occupies almost all of the rest of this chapter of his lecture notes (WS 1916/17 Lectures, 122–156). After a short discussion of the

---

210 "Die mathematischen Schwierigkeiten hindern uns z.B. schon an der Konstruktion eines einzigen neutralen Massenpunktes. Könnten wir eine solche neutrale Masse konstruieren, und würden wir den Verlauf in der Umgebung dieser Stelle kennen, so würden die $g_{\mu\nu}$ wenn wir die neutrale Masse immer mehr gegen einen Massenpunkt hin degenerieren lassen, in diesem Punkte eine Singularität aufweisen. Eine solche müssten wir als erlaubt ansehen in dem Sinne, dass die $g_{\mu\nu}$ ausserhalb der nächsten Umgebung der Singularität den in der Natur wirklich realisierten Verlauf richtig wiedergeben. Eine solche Singularität müssen wir nun in [the Schwarzschild line element] vor uns haben. Im übrigen können wir schon jetzt sagen, dass die Konstruktion eines neutralen Massenpunktes, auch wenn sie später möglich sein wird, sich als so kompliziert erweisen wird, dass man für die Zwecke, in denen man nicht die nächste Umgebung des Massenpunktes betrachtet, mit ausreichender Genauigkeit mit den mit einer Singularität behafteten, angenähert richtigen Gravitationspotentialen wird rechnen können.
Wir behaupten nun Folgendes: Wenn wir die mathematische Entwicklung, die zur Konstruktion eines neutralen Massenteilchens führt, wirklich durchführen können, so werden wir dabei vermutlich auf Gesetze stossen, die wir einstweilen noch axiomatisch formulieren müssen, die aber später sich als Folgen unserer allgemeinen Theorie ergeben werden, als Folgen freilich, die bestimmt nur durch eine weitsichtige Theorie und komplizierte Rechnung zu begründen sein werden. Diese Axiome, die also nur provisorische Geltung haben sollen, fassen wir folgendermassen:
Axiom I: Die Bewegung eines Massenpunktes im Gravitationsfeld wird durch eine geodätische Linie dargestellt, welche eine Zeitlinie ist.
Axiom II: Die Lichtbewegung im Gravitationsfeld wird durch eine geodätische Nulllinie dargestellt.
Axiom III: Eine singuläre Stelle der Massbestimmung ist äquivalent einem Gravitationszentrum." (WS 1916/17 Lectures, 120–121)

dimensions of various physical quantities (WS 1916/17 Lectures, 156–158), he discusses the behavior of measuring threads and clocks in the Schwarzschild gravitational field (WS 1916/17 Lectures, 159–163), and concludes the chapter with a discussion of the third general-relativistic effect treated by Einstein, the gravitational redshift of spectral lines (WS 1916/17 Lectures, 163–166).

In Paper 2, these topics are treated more briefly if at all: Axioms I and II and their motivations, are discussed on pp. 70–71. The discussion of time-like geodesics occupies pp. 71–75, and the paper closes with a discussion of null geodesics on pp. 75–76. In summary, this paper must be considered a singular hybrid between the blossoming of a rich mathematical tradition that Hilbert brings to bear on the problems of general relativity, and the agony of facing the collapse of his own research program.

6.4 Revisions of Paper 2

Paper 2, like Paper 1, was republished twice: Indeed, the two were combined in the 1924 version, Paper 2 becoming Part 2 of *Die Grundlagen der Physik* (Hilbert 1924, 11–32). We shall refer to this version as "Part 2." The reprint of Hilbert 1924 in the *Gesammelte Abhandlungen* was edited by others, presumably under Hilbert's supervision (Hilbert 1935, 268–289). We shall refer to this version as "Part 2–GA." Compared to Paper 1, Hilbert's additions and corrections to Paper 2 are less substantial, as is to be expected since Paper 2 was written largely within the context of general relativity. Most changes are minor improvements, e.g. in connection with recent literature on the theory. There are three significant changes however. One, introduced by Hilbert at the beginning of Part 2, concerns Hilbert's view of the relation between Papers 1 and 2, the other two by the editors of the *Gesammelte Abhandlungen* in Part 2–GA. The second concerns the Cauchy problem, and the third concerns his understanding of invariant assertions. We shall discuss these revisions, both major and minor.

The first significant change concerns the paper's goal: Paper 2 states that "it seems necessary to discuss some more general questions of a logical as well as physical nature" ("erscheint es nötig, einige allgemeinere Fragen sowohl logischer wie physikalischer Natur zu erörtern" Hilbert 1917, 53). Part 2 states: "now the relation of the theory with experience shall be discussed more closely" ("Es soll nun der Zusammenhang der Theorie mit der Erfahrung näher erörtert werden" Hilbert 1924, 11). This revision confirms our interpretation of Paper 2 as resulting, in its original

211 "Prinzipiell aber hat dieses neue Einsteinsche Gesetz gar keine Ähnlichkeit mit dem Newtonschen. Es ist unmöglich komplizierter als das letztere. Wenn wir es trotzdem dem Newtonschen vorziehen, so ist dies darin begründet, dass dieses Gesetz einem tiefliegenden philosophischen Prinzip — dem der allgemeinen Invarianz — genüge leistet, und dass es zwei so heterogene Dinge, wie das Newtonsche Gesetz einerseits und die tatsächliche Gültigkeit der Euklidischen Geometrie in der Physik unter gewissen einfachen Voraussetzungen andererseits als Spezialfälle enthält, sodass wir also nicht, wie dies bis jetzt der Fall war, zuerst die Gültigkeit der Euklidischen Geometrie voraussetzen, und dann ein Attraktionsgesetz anflicken müssen." (WS 1916/17 Lectures, 122)

version, from the tension between Hilbert's concern about the unsolved problems of his theory, in particular the problem of causality, and his immersion in the challenging applications of general relativity, in particular to astronomy. Since Hilbert's revision of Paper 1 had effectively transformed his theory into a version of general relativity, the revision of Paper 2 could now be presented as relating this theory to its empirical basis, the astronomical problems being addressed by contemporary general relativity.

We shall now discuss the changes, which occur in four of the six topics discussed (see above):

1. The metric tensor and its measurement: Part 2 drops all reference to "Messfaden." The discussion of measurement is based entirely on the "Lichtuhr," but is otherwise parallel to that in Paper 2 (Hilbert 1924, 11–13).
2. The causality problem for the field equations (Hilbert 1924, 16–19): There are several changes in the discussion. The wording, with which Hilbert introduces the problem now reads:

   > Our basic equations of physics [the gravitational and the electromagnetic field equations] in no way take the form characterized above [Cauchy normal form]: rather four of them are, as I have shown, a consequence of the rest ...[212]

   Note that "wie ich gezeigt habe" replaces "nach Theorem I" (see p. 59 of Paper 2). Hilbert says that, if there were 4 additional invariant equations, then the system of equations in Gaussian normal coordinates "ein überbestimmtes System bilden würde" (see p. 16 of Part 2) replacing "untereinander in Widerspruch ständen" (see p. 60 of Paper 2).

   In the discussion of the first way, in which "physically meaningful, i.e., invariant assertions can be expressed mathematically" (Hilbert 1917, 62; 1924, 18), he corrects a number of the equations in his example. His discussion of the third way is shortened considerably, now reading:

   > An assertion is also invariant and is therefore always physically meaningful if it is valid for any arbitrary coordinate system, without the need for the expressions occurring in it to possess a formally invariant character.[213]

   In Paper 2, this sentence had ended with "...gültig sein soll," and the paragraph had given the example of Einstein's gravitational energy-momentum complex.
3. Euclidean geometry: His discussion is the same, except that the discussion of gravitational perturbations drops the use of an imaginary time coordinate and Euclidean metric (Hilbert 1924, 19–23, 26).

---

212 "Unsere Grundgleichungen der Physik sind nun keineswegs von der oben charakterisierten Art; vielmehr sind, wie ich gezeigt habe, vier von ihnen eine Folge der übrigen ..." (Hilbert 1924, 16)

213 "Auch ist eine Aussage invariant und hat daher stets physikalischen Sinn, wenn sie für jedes beliebige Koordinatensystem gültig ist, ohne daß dabei die auftretenden Ausdrücke formal invarianten Charakter zu besitzen brauchen." (Hilbert 1924, 19)

4. The Schwarzschild solution (Hilbert 1924, 23–32): He adds a footnote to the light ray axiom:

> Laue has shown for the special case $L = \alpha Q$ [i.e., for the usual Maxwell Lagrangian] how this theorem can be derived from the electrodynamic equations by considering the limiting case of zero wavelength.[214]

followed by a reference to Laue's 1920 paper (Laue 1920) showing that Hilbert kept up with the relativity literature. He also dropped a rather trivial footnote to Axiom I (massive particles follow time-like world lines):

> This last restrictive addition [i.e., "Zeitlinie"] is to be found neither in Einstein nor in Schwarzschild.[215]

He adds a more careful discussion of circular geodesics, the radius of which equals the Schwarzschild radius (Hilbert 1924, 30, compared to 1917, 75), but otherwise the discussion of geodesics remains the same.

When the 1924 version of his two papers was republished in 1935 in his *Gesammelte Abhandlungen*, the editors introduced two extremely significant changes, as well as more trivial ones that we shall not discuss, that retract the last elements of Hilbert's attempt to provide a solution to the causality problem for his theory. These changes in Part 2–GA are footnotes marked "Anm[erkung] d[er] H[erausgeber]". The first occurs in the discussion of the causality principle for generally-covariant field equations (Hilbert 1924, 18–19; 1935, 275–277). The sentence:

> Since the Gaussian coordinate system itself is uniquely determined, therefore also all assertions with respect to these coordinates about those potentials (24) [equation number in the original] are of invariant character.[216]

is dropped; and a lengthy footnote is added (Hilbert 1935, 275–277). This footnote shows that the editors[217] correctly understood the nature of the fourteen field equations. Six of the ten gravitational and three of the four electromagnetic equations contain second time derivatives of the six spatial components of the metric tensor and three spatial components of the electromagnetic potentials. Thus, their values together with those of their first time derivatives on the initial hypersurface determine their evolution off that hypersurface. But these initial values are subject to constraints, set by the remaining four gravitational and one electromagnetic equation, which contain no second time derivative. Due to the differential identities satisfied

214 "Laue hat für den Spezialfall $L = \alpha Q$ [i.e., for the usual Maxwell Lagrangian] gezeigt, wie man diesen Satz aus den elektrodynamischen Gleichungen durch Grenzübergang zur Wellenlänge Null ableiten kann." (Hilbert 1924, 27).

215 "Dieser letzte einschränkende Zusatz findet sich weder bei Einstein noch bei Schwarzschild." (Hilbert 1917, 71)

216 "Da das Gaußische Koordinatensystem selbst eindeutig festgelegt ist, so sind auch alle auf dieses Koordinatensystem bezogenen Aussagen über jene Potentiale (24) von invariantem Charakter." (Hilbert 1924, 18)

217 Paul Bernays, Otto Blumenthal, Ernst Hellinger, Adolf Kratzer, Arnold Schmidt, and Helmut Ulm.

by the field equations, if these constraint equations hold initially, they will continue to hold by virtue of the remaining field equations. This footnote culminates in the statement:

> Thus causal lawfulness does not express the full content of the basic equations; rather, in addition to this lawfulness, these equations also yield *restrictive conditions on the respective initial state*.[218]

The editors also explain that, in the gauge-invariant electromagnetic case, it is only the fields and not the potentials that are determined by the field equations. The editors' addition thus presents a lucid account of the Cauchy problem in general relativity, and shows that Hilbert's attempt to formulate a principle of causality for his theory in terms of the classical notion of initial data (i.e. values that can be freely chosen at any given moment in time, which then determine their future evolution) had not taken into account the existence of constraints on the initial data.

The second footnote occurs in the discussion of how to satisfy the requirement that physically meaningful assertions be invariant by use of an invariant coordinate system (Hilbert 1924, 18–19). The footnote, which actually undermines claims in Hilbert's paper, reads:

> In the case of each of the three types of preferred coordinate systems named here, there is only a partial fixation of the coordinates. The Gaussian nature of a coordinate system is preserved by arbitrary transformations of the space coordinates and by Lorentz transformations, and a coordinate system in which the vector r^k has the components $(0, 0, 0, 1)$, is transformed into another such system by an arbitrary transformation of the spatial coordinates together with a spatially varying shift of the temporal origin.
>
> The characterization of a Gaussian coordinate system by conditions (23) [equation number in the original] and likewise that of the third-named preferred coordinate system through the conditions for r^k is in fact not completely invariant insofar as the specification of the fourth coordinate—introduced through conditions (21) [equation number in the original; the conditions for a "proper" coordinate system]—plays a role in it.[219]

---

218 "Somit bringt die kausale Gesetzlichkeit nicht den vollen Inhalt der Grundgleichungen zum Ausdruck, diese liefern vielmehr außer jener Gesetzlichkeit noch *einschränkende Bedingungen für den jeweiligen Anfangszustand*." (Hilbert 1935, 277)

219 "Bei den drei hier genannten Arten von ausgezeichneten Koordinatensystemen handelt es sich jedesmal nur um eine partielle Festlegung der Koordinaten. Die Eigenschaft des Gaußischen Koordinatensystems bleibt erhalten bei beliebigen Transformationen der Raumkoordinaten und bei Lorentztransformationen, und ein Koordinatensystem, in welchem der Vektor r^k die Komponenten $(0, 0, 0, 1)$ hat, geht wieder in ein solches über bei einer beliebigen Transformation der Raumkoordinaten nebst einer örtlich variablen Verlegung des zeitlichen Nullpunktes.
Die Charakterisierung des Gaußischen Koordinatensystems durch die Bedingungen (23) und ebenso die des drittgenannten ausgezeichneten Koordinatensystems durch die Bedingungen für r^k ist übrigens insofern nicht völlig invariant, als darin die Auszeichnung der vierten Koordinate zur Geltung kommt, die mit der Aufstellung der Bedingungen (21) eingeführt wurde." (Hilbert 1935, 277)

The editors of Hilbert's papers corrected two major mathematical errors that survived his own revision of Paper 2, and since he was still active when this edition of his papers was published, it can be assumed that these changes were made with his consent, if not participation.

7. THE FADING AWAY OF HILBERT'S POINT OF VIEW AND ITS SUBSUMPTION BY EINSTEIN'S PROGRAM

Early on, Einstein and Weyl set the tone for the way in which Hilbert's papers on the *Foundations of Physics* were integrated into the mainstream of research in physics and mathematics. Not only did the articles by Einstein and Weyl receive immediate attention when first published in the *Sitzungsberichte* of the Prussian Academy of Sciences, but they were soon incorporated into successive editions of *Das Relativitätsprinzip*, then the standard collection of original works on the development of relativity.[220] Three out of four of Einstein's works added to the third edition mention Hilbert, as does Weyl's contribution to the fourth edition—although, as we shall see, the latter's omissions are as significant as his attributions. Translated into French, English and other languages, and in print to this day, countless scholars had their impression of the scope and history of relativity shaped by this book.

First we shall discuss Einstein's two mentions of Hilbert in 1916. (His third in 1919 is related to Weyl's 1918 paper, so we shall discuss it afterwards.) In contrast with Hilbert's need to reorganize his theory in reaction to Einstein's work, Einstein could assimilate Hilbert's results into the framework of general relativity without being bothered by the latter's differing interpretation of them. This assimilation, in turn, assigned Hilbert a place in the history of general relativity.

Einstein's 1916 review paper on general relativity mentions Hilbert in a discussion of the relation between the conservation identities for the gravitational field equations and the field equations for matter:

> Thus the field equations of gravitation contain four conditions [the conservation equations for the energy-momentum tensor of matter] which govern the course of material phenomena. They give the equations of material processes completely of the latter are capable of being characterized by four independent differential equations.[221]

220 See (Blumenthal 1913; 1919; 1923; 1974). All editions were edited by the mathematician Otto Blumenthal. The first edition appeared as the second volume of his series *Fortschritte der Mathematischen Wissenschaften in Monographien* (the first being a collection of Minkowski's papers on electrodynamics), "als eine Sammlung von Urkunden zur Geschichte des Relativitätsprinzips" ("Vorwort" [n.p.]). The third edition in 1919 included additional papers by Einstein on general relativity, the fourth edition added Weyl's first paper on his unified theory of gravitation and electromagnetism. The fifth edition in 1923 is the basis of the editions currently in print, and of the translations into other languages. It would be interesting to know how Blumenthal chose the papers to include in what became the canonical source book on relativity.

A footnote adds a reference to Paper 1.[222] Thus, Einstein subsumed into the general theory of relativity, as a particular case of an important general result, what Hilbert regarded as an outstanding achievement of his theory. Hilbert's interpretation of this result as embodying a unique coupling between gravitation and electromagnetism, is not even mentioned.

In the same year, Einstein published his own derivation of the generally-covariant gravitational field equations from a variational principle. While in the 1916 review paper he had given a non-invariant "Hamiltonian" (= Lagrangian) for the field equations modulo the coordinate condition $\sqrt{-g} = 1$, he now proceeded in a manner reminiscent of Hilbert's in Paper 1. He uses the same gravitational variables (the $g_{\mu\nu}$ and their first and second derivatives), but Einstein's $q_{(\rho)}$ "describe matter (including the electromagnetic field" ("beschreiben die Materie (inklusive elektromagnetisches Feld)") and hence are arbitrary in number and have unspecified tensorial transformation properties. By his straightforward generalization, Einstein transformed Hilbert's variational derivation into a contribution to general relativity, without adopting the latter's perspective on this derivation as providing a synthesis between gravitation and a specific theory of matter. Rather, Einstein's generalization made it possible to regard Hilbert's theory as no more than a special case.

Einstein prefaced his calculations with some observations placing his work in context:

> H. A. Lorentz and D. Hilbert have recently succeeded [footnoted references to Lorentz's four papers of 1915–1916 and Hilbert's Paper 1] in presenting the theory of general relativity in a particularly comprehensive form by deriving its equations from a single variational principle. The same shall be done in this paper. My aim here is to present the fundamental connections as transparently and comprehensively as the principle of general relativity allows. In contrast to Hilbert's presentation, I shall make as few assumptions about the constitution of matter as possible.[223]

---

221 "The Foundation of the General Theory of Relativity" p. 810, in (CPAE 6E, Doc. 30, 187). "Die Feldgleichungen der Gravitation enthalten also gleichzeitig vier Bedingungen [the conservation equations for the energy-momentum tensor of matter], welchen der materielle Vorgang zu genügen hat. Sie liefern die Gleichungen des materiellen Vorganges vollständig, wenn letzterer durch vier voneinander unabhängige Differentialgleichungen charakterisierbar ist." (Einstein 1916a, 810)

222 The reference to "p. 3," is probably to a separately paginated off-print; see the discussion in (Sauer 1999).

223 Einstein, "Hamilton's Principle and the General Theory of Relativity" Sitzungsberichte 1916, 1111–1116, citation from p. 1111, in (CPAE 6E, Doc. 41, 240). "In letzter Zeit ist es H. A. Lorentz und D. Hilbert gelungen [footnoted references to Lorentz's four papers of 1915–1916 and Hilbert's Paper 1], der allgemeinen Relativitätstheorie dadurch eine besonders übersichtliche Gestalt zu geben, daß sie deren Gleichungen aus einem einzigen Variationsprinzipe ableiteten. Dies soll auch in der nachfolgenden Abhandlung geschehen. Dabei ist es mein Ziel, die fundamentalen Zusammenhänge möglichst durchsichtig und so allgemein darzustellen, als es der Gesichtspunkt der allgemeinen Relativität zuläßt. Insbesondere sollen über die Konstitution der Materie möglichst wenig spezialisierende Annahmen gemacht werden, im Gegensatz besonders zur Hilbertschen Darstellung." (Einstein 1916b, 1111)

Thus Einstein both gave Hilbert credit for his accomplishments and circumscribed their nature: Like Lorentz, Hilbert was supposedly looking for a variational derivation of the general-relativistic field equations, but included assumptions about the constitution of matter that were too special. In an earlier, unpublished draft, Einstein's tone was even sharper:

> Hilbert, following the assumption introduced by Mie that the H function depends on the components of a four-vector and their first derivatives, I do not consider very promising.[224]

In private correspondence, he was still more harsh, but also gave his reasons for disregarding Hilbert's point of view:

> Hilbert's assumption about matter appears childish to me, in the sense of a child who knows none of the perfidy of the world outside. [...] At all events, mixing the solid considerations originating from the relativity postulate with such bold, unfounded hypotheses about the structure of the electron or matter cannot be sanctioned. I gladly admit that the search for a suitable hypothesis, or Hamilton function, for the construction of the electron, is one of the most important tasks of theory today. The "axiomatic method" can be of little use here, though.[225]

Evidently, Einstein clearly perceived the diverse status of the physical assumptions underlying general relativity, on the one hand, and Hilbert's theory, on the other. From Einstein's point of view, Hilbert's detailed results, such as his variational derivation of the Schwarschild metric could be—and were—acknowledged as contributions to the development of general relativity, without any need to refer to the grandiose program, within which Hilbert had originally placed them.

In view of his own claims in this regard, one might expect Hilbert's work to have played a prominent role in the developing search for a unified field theory.[226] But his fate was that of a transitional figure, eclipsed by both his predecessors and his successors. His achievements were perceived as individual contributions to general relativity rather than as genuine milestones on the way towards a unified field theory. Evidently, this "mixed score" was the price Hilbert had to pay for being made one of the founding fathers of general relativity.

In his first contribution to unified field theory, Weyl assigned a definite place to Hilbert, if largely by omission. After presenting his generalization of Riemannian

---

224 "Die von Hilbert im Anschluss an Mie eingeführte Voraussetzung, dass sich die Funktion H durch die Komponenten eines Viererverktors q_ρ und dessen erste Ableitungen darstellen lasse, halte ich für wenig aussichtsvoll." See note 3 to Doc. 31 in (CPAE 6, 346).

225 "Der Hilbertsche Ansatz für die Materie erscheint mir kindlich, im Sinne des Kindes, das keine Tükken der Aussenwelt kennt. [...] Jedenfalls ist es nicht zu billigen, wenn die soliden Überlegungen, die aus dem Relativitätspostulat stammen, mit so gewagten, unbegründeten Hypothesen über den Bau des Elektrons bezw. der Materie verquickt werden. Gerne gestehe ich, dass das Aufsuchen der *geeigneten* Hypothese bezw. Hamilton'schen Funktion für die Konstruktion des Elektrons eine der wichtigsten heutigen Aufgaben der Theorie bildet. Aber die "axiomatische Methode" kann dabei wenig nützen." Einstein to Hermann Weyl, 23 November 1916, (CPAE 8, 365–366).

226 For a historical discussion, see (Majer and Sauer 2005; Goenner 2004).

geometry to include what he called "gauge invariance" (Eichinvarianz),[227] Weyl turned to unified field theory:

> Making the transition from geometry to physics, we must assume, in accord with the example of Mie's theory [references to Mie's papers of 1912/13 and Weyl's recently-published *Raum-Zeit-Materie*], that the entire lawfulness of nature is based upon a certain integral invariant, the action
>
> $$\int W d\omega = \int \mathfrak{W} dx \qquad (\mathfrak{W} = W\sqrt{g}),$$
>
> in such a way that the actual world is distinguished from all possible four-dimensional metric spaces, by the fact that the action contained in every region of the world takes an extremal value with respect to those variations of the potentials g_{ij}, ϕ_i that vanish at the boundaries of the region in question.[228]

In spite of its obvious relevance, there is no mention here of Hilbert. The sole mention comes in what we shall refer to as "the litany" since this or a similar list occurs so frequently in the subsequent literature:

> We shall show in fact, in the same way that, according to the investigations of Hilbert, Lorentz, Einstein, Klein and the author [reference follows to Paper 1 for Hilbert], the four conservation laws of matter (of the energy-momentum-tensor) are connected with the invariance of the action under coordinate transformations containing four arbitrary functions; the charge conservation law is linked to a newly introduced "scale-invariance" depending on a fifth arbitrary function.[229]

This passage, (incorrectly) attributing to Hilbert a clarification of energy-momentum conservation in general relativity and disregarding his attempt to create a unified field theory, makes his "mixed score" particularly evident. In a footnote added to the republication of his paper in *Das Relativitätsprinzip,* Weyl notes that:

> The problem of defining all invariants W admissible as actions, while requiring that they contain the derivatives of g_{ij} up to second order at most, and those of ϕ_i only up to first order, was solved by R. Weitzenböck [Weitzenböck 1920], [230]

---

227 This generalization was named a Weyl space by J.A. Schouten (see Schouten 1924).

228 "Von der Geometrie zur Physik übergehend, haben wir nach dem Vorbild der Mieschen Theorie anzunehmen, daß die gesamte Gesetzmäßigkeit der Natur auf einer bestimmten Integralinvariante, der Wirkungsgröße $\int W d\omega = \int \mathfrak{W} dx \qquad (\mathfrak{W} = W\sqrt{g})$ beruht, derart, daß die wirkliche Welt unter allen möglichen vierdimensionalen metrischen Räumen dadurch ausgezeichnet ist, daß für sie die in jedem Weltgebiet enthaltene Wirkungsgröße einen extremalen Wert annimmt gegenüber solchen Variationen der Potentiale g_{ik}, ϕ_i, welche an den Grenzen des betreffenden Weltgebiets verschwinden." (Weyl 1918c, 475)

229 "Wir werden nämlich zeigen: in der gleichen Weise, wie nach Untersuchungen von Hilbert, Lorentz, Einstein, Klein und dem Verf. [reference follows to Paper 1 for Hilbert] die vier Erhaltungsätze der Materie (des Energie-Impuls-Tensors) mit der, vier willkürliche Funktionen enthaltenden Invarianz der Wirkungsgröße gegen Koordinatentransformationen zusammenhängen, ist mit der hier neu hinzutretenden, eine fünfte willkürliche Funktion hereinbringenden "Maßstab-Invarianz" [...] das Gesetz von der Erhaltung der Elektrizität verbunden." (Weyl 1918c, 475)

without mentioning that this is the solution to the problem raised by Hilbert's ansatz for the invariant Lagrangian, first introduced in Paper 1. Little wonder that those whose knowledge of the history of relativity came from *Das Relativitätsprinzip* had no idea of Hilbert's original aims and little more of his achievements.

Hilbert fared a little better in Weyl's *Raum-Zeit-Materie,* the first treatise on general relativity (Weyl 1918a; 1918b; 1919; 1921; 1923).[231] The discussion of the energy-momentum tensor in the first edition (section 27) credits Hilbert with having shown that (Weyl 1918a; 1918b, 184):

> [...] Mie's electrodynamics can be generalized from the assumptions of the special to those of the general theory of relativity. This was done by Hilbert.[232]

Footnote 5 cites Paper 1 and adds (Weyl 1918a; 1918b, 230):

> The connection between Hamilton's function and the energy-momentum tensor is established here, and the gravitational equations articulated almost simultaneously with Einstein, if only within the confines of Mie's theory,[233]

Hilbert's work has already been subsumed under general relativity. Curiously, both textual reference and footnote disappear from all later editions (but see the discussion below of the fifth edition). Presumably because Weyl had already mentioned Hilbert, the latter's name does not appear in the litany in the first edition (footnote 6), listing those who had worked on the derivation of the energy-momentum conservation laws. By the third edition, Hilbert has been added to the litany (Weyl 1919, 266 n. 8), and remained there. In his discussion of causality for generally-covariant field equations in the first edition, Weyl credits Papers I and II (Weyl 1918a; 1918b, 190 and 230, n. 9); again, this note disappears from all later editions. Paper 2 is also cited in the first edition in connection with the Schwarzschild solution (Weyl 1918a; 1918b, 230, n. 15), and the introduction of geodesic normal coordinates (Weyl 1918a; 1918b, 230, n. 21).

The third edition carries over these references to Paper 2 and adds one in connection with linearized gravitational waves (Weyl 1919, 266, n. 14); and the fourth edition includes all these footnotes. Perhaps questions had been raised concerning

---

230 "Die Aufgabe, alle als Wirkungsgrößen zulässigen invarianten W zu bestimmen, wenn gefordert ist, daß sie die Ableitungen der g_{ik} höchstens bis zur 2., die der ϕ_i nur bis zur 1. Ordnung enthalten dürfen, wurde von R. Weitzenböck [Weitzenböck 1920] gelöst." (Blumenthal 1974, 159; translation from Lorentz et al. 1923.) This seventh edition from 1974 is an unchanged reprint of the fifth edition of 1923, 159, n. 2. Weitzenböck has his own version of the litany: "Die obersten physikalischen Gesetze: Feldgesetze und Erhaltungsätze werden nach den klassischen Arbeiten von Mie, Hilbert, Einstein, Klein und Weyl aus einem Variationsprinzip [...] hergeleitet"(p. 683). It is not clear why Lorentz is omitted from the litany; perhaps he was too much of a physicist for Weitzenböck.

231 The second edition of 1918 was unchanged, the fourth of 1921 was translated into English and French; the fifth of 1923, being thereafter reprinted without change.

232 "[...] die Miesche Elektrodynamik von den Voraussetzungen der speziellen auf die der allgemeinen Relativitätstheorie übertragen werden [kann]. Dies ist von Hilbert durchgeführt worden."

233 "Hier ist auch der Zusammenhang zwischen Hamiltonscher Funktion and Energie-Impuls-Tensor aufgestellt und wurden, etwa gleichzeitig mit Einstein, wenn auch nur im Rahmen der Mieschen Theorie, die Gravitationsgleichungen ausgesprochen."

Weyl's treatment of Hilbert in the book; at any rate, the footnote to the litany citing Hilbert in the fifth edition again credits him with a contribution to general relativity, rather than to unified field theories:

> In the first communication, Hilbert established the invariant field equations simultaneously with and independently of Einstein, but within the framework of Mie's hypothetical theory of matter.[234]

In short, in none of the editions is Hilbert mentioned in connection with unified field theories.

Pauli's standard 1921 review article on relativity is another major source, still consulted mainly in the English translation of 1958 (with additional notes) by physicists and mathematicians for historical and technical information about relativity and unified field theories (Pauli 1921; 1958). Pauli adopted what we may call the Einstein-Weyl line on Hilbert, considering him a somewhat unfortunate founding father of general relativity. After describing Einstein's work on general relativity culminating in the November 1915 breakthrough, Pauli adds in a footnote (Pauli 1921):[235]

> At the same time as Einstein, and independently, Hilbert formulated the generally covariant field equations [reference to Paper 1]. His presentation, though, would not seem to be acceptable to physicists, for two reasons. First, the existence of a variational principle is introduced as an axiom. Secondly, of more importance, the field equations are not derived for an arbitrary system of matter, but are specifically based on Mie's theory of matter ... .

His discussion of invariant variational principles in section 23 cites the litany: "investigations by Lorentz, Hilbert, Einstein, Weyl and Klein[236] on the role of Hamilton's Principle in the general theory of relativity" (Pauli 1921).[237]

Later (section 56), he discusses the question of causality in "a generally relativistic [i.e, generally-covariant] theory," arguing from general covariance to the existence of 4 identities between the 10 field equations, and concluding (Pauli 1921):[238]

> The contradiction with the causality principle is only apparent, since the many possible solutions of the field equations are only formally different. Physically they are completely equivalent. The situation described here was first recognized by Hilbert.

This passage represents a striking example of erroneously crediting Hilbert with a contribution to general relativity while neglecting his actual achievements. To make matters worse, Pauli's footnote cites Paper 1, rather than Paper 2; after also crediting Mach with a version of this insight, he adds (Pauli 1921):[239]

---

234 "In der 1. Mitteilung stellte Hilbert gleichzeitig und unabhängig von Einstein die invarianten Feldgleichungen auf, aber im Rahmen der hypothetischen Mieschen Theorie der Materie." (Weyl 1923, 329, n. 10)

235 Section 50, cited from translation in (Pauli 1958, 145 n. 277).

236 See Felix Klein to Wolfgang Pauli, 8 May 1921 in (Pauli 1979, 31).

237 Cited from translation in (Pauli 1958, 68).

238 Cited from translation in (Pauli 1958, 160).

239 Cited from translation in (Pauli 1958, 160, n. 315).

> Furthermore it deserves mentioning that Einstein had, for a time, held the erroneous view that one could deduce from the non-uniqueness of the solution that the gravitational equations could not be generally-covariant [reference to (Einstein 1914b)].

Pauli does acknowledge various contributions to general relativity in Paper 2.[240] But his discussion of unified field theories (Part V), like Weyl's, jumps from Mie (section 64) to Weyl (section 65) without mention of Hilbert.

By examining a couple of early treatises on relativity by non-German authors, we can get some idea of the propagation of the Einstein-Weyl line as canonized by Pauli. Jean Becquerel's *Le Principe de la Relativité et la Théorie de la Gravitation* was the first French treatise on general relativity. In Chapter 16 on "Le Principe d'Action Stationnaire," Becquerel asserts:

> Lorentz and Hilbert [references to Papers 1 and 2], and then Einstein succeeded in presenting the general equations of the theory of gravitation as consequences of a unique stationary action principle, …[241]

followed by section 103 on "Méthode de Lorentz et d'Hilbert" (Becquerel 1922, 257–262). Paper 2 is cited in connection with linearized gravitational waves (Becquerel 1922, 216), but there is no mention of Hilbert in Chapter 18 on "Union du Champ de Gravitation et du Champ Électromagnétique. Géometries de Weyl et d'Eddington" (Becquerel 1922, 309–335).

Until recently Eddington's treatise, *The Mathematical Theory of Relativity,* was widely read, cited and studied by students; and was translated into French and German (Eddington 1923; 1924). The two English editions cite Papers 1 and 2 in the bibliography, with a reference to section 61 on "A Property of Invariants,"[242] which demonstrates the theorem:[243]

> The Hamiltonian [i.e, Lagrangian] derivative of any fundamental invariant is a tensor whose divergence vanishes.

Outside the Bibliography, few references are given in the English editions; but Eddington added material to the German translation, including several references to Hilbert (Eddington 1925). On p. 114, footnote 1 credits Hilbert (Paper 2) with realizing that the assumption of asymptotic flatness is not needed in the derivation of the Schwarzschild metric. On p. 116, he credits Paper 2 for an "elegante Methode" for deducing the Christoffel symbols from the geodesic equation; and on p. 183, he credits the same paper for the first strict proof that one can always satisfy the linearized

240 See (Pauli 1921), section 13 for Axiom II; section 22 for discussion of the restrictions on coordinate systems if three coordinates are to be space-like and one time-like; and section 60 for the proof that linearized harmonic coordinate conditions may always be imposed.

241 "Lorentz et Hilbert [references to Papers 1 and 2], puis Einstein, ont réussi à presenter les équations générales de la theorie de la gravitation comme des conséquences d'un unique principe d'action stationnaire," (Becquerel 1922, 256).

242 See (Eddington 1924, 264): "wherever possible the subject matter is indicated by references to the sections in this book chiefly concerned."

243 See (Eddington 1924, 140–141).

harmonic coordinate conditions by an infinitesimal coordinate transformation. And that is it.

We see that, by the mid-1920s, and with minor variations within the accepted limits, the Einstein-Weyl line on Hilbert's role was already becoming standard in the literature on relativity.

8. AT THE END OF A ROYAL ROAD

The preceding discussion has shown that Hilbert did not discover a royal road to the field equations of general relativity. In fact, he did not formulate these equations at all but, at the end of 1915, developed a theory of gravitation and electromagnetism that is incompatible with Einstein's general relativity. Nevertheless, this theory can hardly be considered an achievement parallel to that of Einstein's creation of general relativity, to be judged by criteria independent of it. Not only is the dependence of Hilbert's theory on and similarity to Einstein's earlier, non-covariant Entwurf theory of gravitation too striking; but its contemporary reception as a contribution to general relativity and regardless of the extent to which Hilbert accepted the transformation of his theory into such a contribution, this is evidence of the theory's evanescent and heteronomous character. It could thus appear as if our account, in the end, describes a race for the formulation of a relativistic theory of gravitation with a clear winner—Einstein—and a clear loser—Hilbert. In contrast to the legend of Hilbert's royal road, such an account would bring us essentially back to Pauli's sober assessment of Hilbert's work as coming close to the formulation of general relativity but being faulted by its dependence on a specific theory of matter. However, as we have shown, this interpretation ascribes to Hilbert results in general relativity that he neither intended nor achieved, and ignores contributions that lay outside the scope of general relativity but were nevertheless crucial for its development. In view of such conundrums, we therefore propose not to consider the Einstein-Hilbert race as the competition between two individuals and their theories but as an event within a larger, collective process of knowledge integration.

As formulated by Einstein in 1915, general relativity incorporates elements of classical mechanics, electrodynamics, the special theory of relativity, and planetary astronomy, as well as such mathematical traditions as non-Euclidean geometry and the absolute differential calculus. It integrates these elements into a single, coherent conceptual framework centered around new concepts of space, time, inertia and gravitation. Without this enormous body of knowledge as its underpinning, it would be hard to explain the theory's impressive stability and powerful role even in today's physics. This integration was the result of an extended and conflict-laden process, to which not only Einstein but many other scientists contributed. From the point of view of historical epistemology, it was a collective process in an even deeper sense.[244] It involved a substantial, shared knowledge base, structured by fundamental concepts, models, heuristic etc., which were transmitted by social institutions, utilizing material representations, such as textbooks, and appropriated by individual learning pro-

cesses. While individual thinking is governed to a large degree by these shared resources, it also affects and amplifies them, occasionally even changing these epistemic structures. On the basis of such an epistemology, which takes into account the interplay between shared knowledge resources and individual thinking, the emergence and fading away of a theory such as Hilbert's can be understood as an aspect of the process of integration of knowledge that produced general relativity.

To answer the question of from where alternative solutions (or attempted solutions) to the same problem come, we shall look at some of the shared knowledge of the time available for formulating theories such as those of Einstein and Hilbert. To explain the fading-away of Hilbert's theory, we then discuss the interplay between individual thinking and the knowledge resources that led to the formulation of general relativity and the transformation of Hilbert's theory into a contribution to it. It will become clear that, in both cases, the same mechanism was at work. In the case of general relativity, it integrated the various components of shared knowledge and resulted in the creation of a stable epistemic structure, which represents that integrated knowledge. In the case of Hilbert's theory, the same process disaggregated the various components of shared knowledge that had been brought together in a temporary structure, and rearranged and integrated them into a more stable structure.

The available knowledge offered a limited number of approaches to the problem that occupied both Einstein and Hilbert in late 1915: the formulation of differential equations governing the inertio-gravitational potential represented by the metric tensor. Two fundamentally different models underlying contemporary field theories of electrodynamics embodied the principal alternatives. One, the "monistic model," conceived all physical phenomena, including matter, in terms of fields. The other "fields-with-matter-as-source model" (or "Lorentz model") was based on a dualism of fields and matter. The first model was the basis for attempts to formulate an "electromagnetic world picture," which remained fragmentary and never succeeded in accounting for most contemporary physical knowledge. The second model was the basis for Lorentz's formulation of electron theory, the epitome of classical electrodynamics, in which matter acts as source for electrodynamic fields that, in turn, affect the motion of material bodies. Rather than attempting to reduce classical mechanical concepts to electrodynamic field concepts, the task associated with the electrodynamic world picture, Lorentz's electron theory successfully integrated electromag-

244 See (Csikszentmihalyi 1988): "All of the definitions [of creativity] ... of which I am aware assume that the phenomenon exists... either inside the person or in the work produced... After studying creativity for almost a quarter of a century, I have come to the reluctant conclusion that this is not the case. We cannot study creativity by isolating individuals and their works from the social and historical milieu in which their actions are carried out. This is because what we call creative is never the result of individual actions alone; it is the product of three main shaping forces: a set of social institutions or *field*, that selects from the variations produced by individuals those that are worth preserving; a stable cultural *domain* that will preserve and transmit the selected new ideas or forms to the following generations; and finally the *individual*, who brings about some change in the domain, a change that the field will consider to be creative." This concept is further discussed in (Stachel 1994).

netic and classical mechanical phenomena. The first model became the core of Hilbert's approach in an attempt to create a unified field theory, while Einstein's search for gravitational field equations was guided by the second. To a large extent, the difference between the two models accounts for the differences between Hilbert's and Einstein's approaches, including their differing capacity to incorporate available physical knowledge into their theories. The information about matter compatible with Hilbert's theory was essentially only Mie's speculative theory: The source-term in Einstein's gravitational field equations could embody the vast amount of information contained in special-relativistic continuum theory, including energy-momentum conservation, as well as Maxwell's theory.

The information available for solving the problem of gravitation was not exhausted by the two different physical models of the interaction between fields and matter. Contemporary mathematics also provided a reservoir of useful tools. The series of attempts between 1912 and 1915 to formulate a theory of gravitation, including contributions by Abraham, Nordström, and Mie, as well as Einstein and Hilbert, illustrates the range of mathematical formalisms available, from partial differential equations for a scalar field to the absolute differential calculus applied to the metric tensor. As did the physical models, different mathematical formalisms showed varying capacities for integrating the available knowledge about matter and gravitation, such as that embodied in Newtonian gravitation theory or in the observational results on Mercury's perihelion shift. To explore its capacity to integrate knowledge, a formalism needs to be elaborated and its consequences interpreted, if possible, as representations of that knowledge. The degree of such successful elaboration and interpretation, the "exploration depth" of a given formalism, determines its acceptability as a possible solution to the physical problem at hand. In early 1913, believing that the Newtonian limit could not be recovered from generally-covariant field equations, Einstein proposed the non-covariant *Entwurf* theory, from which it could be. At the end of 1915, on the basis of an increased "exploration depth" of the formalism, he decided in favor of generally-covariant equations.

Which physical models and mathematical formalisms are favored in a given historical situation depends on many factors, among them their accessibility and specific epistemological preferences that make some of them appear more attractive to certain groups than others. It was natural for a mathematician of Hilbert's caliber to start from a generally-covariant variational principle based on the metric tensor, while Einstein, ignorant of the appropriate mathematical resources, initially tried to develop his own, "pedestrian" calculus for dealing with the metric tensor.[245] It is clear that the monistic field theory model must have appealed more to Hilbert, a mathematician in search for an axiomatic foundation for all of physics, than the conceptually more clumsy dualistic model. The latter, on the other hand, was a more natural starting point for physicists such as Abraham, Einstein, and Nordström, who were familiar

245 See his calculations in "Einstein's Zurich Notebook," e.g. on p. 08L (in vol. 1 of this series). See also the "Commentary" (in vol. 2).

with the extraordinary successes of this model in the domain of electromagnetism. Images of knowledge also determine decisions on the depth and direction of exploration of a given formalism. While the question of the Newtonian limit was crucial to the physicist Einstein, Hilbert did not deal at all with this problem.

Constructs formulated by individual scientists, such as Hilbert's proposal for an axiomatic foundation of physics, are largely contingent; but their building blocks (concepts, models, techniques) are taken from the reservoir of the socially available knowledge characteristic of a given historical situation. This reservoir of shared background knowledge accounts for more than just the intercommunicability of individual contributions such as those of Hilbert and Einstein. Given that such contributions are integrated into already-shared knowledge by various processes of intellectual communication and assimilation, an equilibration process must take place between the individual constructs and the shared knowledge-reservoir. It is the outcome of this process that decides on whether a research program is progressive or degenerating in the sense of Lakatos but also the fate of an individual contribution, its longevity (the case of general relativity), its mutation, or its rapid fading-away (the case of Hilbert's contribution).

Whatever is individually constructed will be brought into contact with other elements of the shared knowledge-base, and thus integrated into it in multiple ways that, of course, are shaped by the social structures of scientific communication. The fate of an individual construct depends on the establishment of such connections. If individual constructs are not embedded, for whatever reasons, within the structures of socially available knowledge, they effectively disappear; if they are so embedded, they will be transmitted as part of shared knowledge. Usually, individual contributions are not assimilated wholesale to shared knowledge but only in a piecemeal fashion. One finds Hilbert's name associated, for instance, with the variational derivation of the field-equations but not with the program of an axiomatic foundation of physics. The "packaging" of individual contributions as they are eventually transmitted and received by a scientific community is not governed by the individual perspectives of their authors but by the more stable cognitive structures of the shared knowledge. The reception of Hilbert's contribution is thus not different from that of most scientific contributions that become assimilated into the great banquet of shared knowledge. It rarely happens that its basic epistemic structures, such as the concepts of space and time in classical physics, are themselves challenged by the growth of knowledge. Usually, these fundamental structures simply overpower any impact of individual contributions by the sheer mass of integrated knowledge they reflect. Only when individual constructs come with their own power of integrating large chunks of shared knowledge do they have a chance of altering these structures. This, in turn, only happens when the individual contributions themselves result from a process of knowledge integration and its reflection in terms of new epistemic structures.

Einstein's theory of general relativity is the result of such an integration process. Over a period of several years, he had attempted not only to reconcile classical physical knowledge about gravitation with the special-relativistic requirement of the finite propagation speed of physical interactions; but also with insights into the inseparabil-

ity of gravitation and inertia, and with the special-relativistic generalization of energy-momentum conservation. Each of these building blocks: Newtonian theory, metric structure of space and time, the equivalence principle, and energy-momentum conservation, was associated with a set of possible mathematical representations, more or less well defined by physical requirements. In the case of energy-momentum conservation, for instance, Einstein had quickly arrived at an appropriate mathematical formulation, which stayed fixed throughout his search for the gravitational field equations. The inseparability of gravitation and inertia as expressed by the equivalence principle, on the other hand, could be given various mathematical representations; for Einstein the most natural at the time seemed to be the role of the metric tensor as the potentials for the inertio-gravitational field. The available mathematical representations of Einstein's building blocks were not obviously compatible with each other. In order to develop a theory comprising as much as possible of the knowledge incorporated in these building blocks, Einstein followed a double strategy.[246] On the one hand, he started from those physical principles that embody the vast store of knowledge in classical and special-relativistic physics and explored the consequences of their mathematical representations in terms of the direction of his other building blocks (his "physical strategy"). On the other hand, he started from those building blocks that had not yet been integrated into a physical theory, such as his equivalence principle, chose a mathematical representation, and explored its consequences, in the hope of being able to find a physical interpretation that also would integrate his other building blocks (his "mathematical strategy"). Eventually, he succeeded in formulating a theory that complies with these heterogeneous requirements; but only at the price of having to modify, in a process of reflection on his own premises, some of the original building blocks themselves, with far-going consequences for the structuring of the physical knowledge embodied in these building blocks, e.g. about the meaning of coordinate systems in a physical theory. That such modifications eventually became more than just personal idiosyncrasies and have had a lasting effect on the epistemic structures of physical knowledge is due to the fact that they were stabilized by the knowledge they helped to integrate into general relativity.

Hilbert's theory was clearly not based on a comparable process of knowledge integration and hence shared the fate of most scientific contribution: dissolution and assimilation to the structures of shared knowledge. Even if, in 1915, he had derived the field equations of general relativity, his theory would not have had the same "exploration depth" as that of Einstein's 1915 version, and hence not covered a similarly large domain of knowledge. Hilbert's theory is rather comparable to one of Einstein's early intermediate versions, for instance to that involving the (linearized) Einstein tensor, briefly considered in the Zurich Notebook in the winter of 1912/13. Einstein quickly rejected this candidate because it appeared to him impossible to derive the Newtonian limit from it, while Hilbert intended to publish his version in late 1915, although he had not checked its compatibility with the Newtonian limit.

246 See "Pathways out of Classical Physics …" (in vol. 1 of this series).

This difference in reacting to a similar candidates for solving the problem of the gravitational field equations obviously does not reveal any difference in the epistemic status of Hilbert's theory compared to Einstein's intermediate version but only by a different attitude with regard to a given exploration depth, motivated by the different image of knowledge that Hilbert associated with his endeavor. Such motivations make little difference to the fate of a theory in the life of the scientific community. In fact, the subsequent elaborations, revisions, and transformations of Hilbert's result testify to an equilibration process similar to that also undergone by Einstein's intermediate versions, in which ever new elements of shared knowledge found their way into Hilbert's construct. In the end, as we have seen, his theory comprises the same major building blocks of physical knowledge as those, on which general relativity is based. The exchange with Einstein and others had effectively compensated for Hilbert's original neglect of the need to consider his results in the light of physical knowledge, and thus substituted, in a way, for the "physical strategy" of Einstein's heuristics, constituting a "collective process of reflection." The fact that the equilibration process leading to general relativity essentially went on in private exchanges between Einstein and a few collaborators, while the equilibration process transforming Hilbert's theory of everything into a constituent of general relativity went on in public, as a contest between Einstein and Hilbert, Berlin and Göttingen, physics and mathematics communities, plays an astonishingly small role in the history of knowledge.

9. ACKNOWLEDGEMENTS

The research for this paper was conducted as part of a collaborative research project on the history of general relativity at the Max Planck Institute for the History of Science. We would like to warmly thank our colleagues in this project, in particular Leo Corry, Tilman Sauer, and Matthias Schemmel, for the discussions, support, and criticism, all of it intensive, which have made this collaboration a memorable and fruitful experience for both of us. Without the careful editorial work by Stefan Hajduk, who not only checked references, consistency, and language but also coordinated the contributions of the authors when they were not in one place, the paper would not have reached its present form. The authors are furthermore indebted to Giuseppe Castagnetti, Peter Damerow, Hubert Goenner, Michel Janssen, Ulrich Majer, John Norton, and David Rowe for helpful comments and conversations on the subject of this paper. We are particularly grateful to the library staff of the Max Planck Institute for the History of Science and its head, Urs Schoepflin, for their creative and indefatigable support in all library-related activities of our work. We thank the Niedersächsische Staats- und Universitätsbibliothek Göttingen (*Handschriftenabteilung*) and the library of the Mathematisches Institut, Universität Göttingen, for making unpublished material available to our project. Finally, we thank Ze'ev Rosenkranz and the Einstein Archives, The Hebrew University of Jerusalem, for permission to quote from Einstein's letters.

REFERENCES

Belinfante, F. J. 1939. "Spin of Mesons." *Physica* 6:887–898.

Becquerel, Henri. 1922. *Le Principe de la Relativité et la Théorie de la Gravitation*. Paris: Gauthier-Villars.

Birkhoff, George D., and Rudolph E. Langer. 1923. *Relativity and Modern Physics*. Cambridge, Ma.: Harvard University Press.

Blumenthal, Otto, ed. 1913. *Das Relativitätsprinzip*. 1st ed. Leipzig, Berlin: Teubner.

———. 1919. *Das Relativitätsprinzip*. 3rd ed. Leipzig, Berlin: Teubner.

———. 1923. *Das Relativitätsprinzip*. 5th ed. Leipzig, Berlin: Teubner.

———. 1974. *Das Relativitätsprinzip*. 7th ed. Darmstadt: Wissenschaftliche Buchgesellschaft.

Born, Max. 1914. "Der Impuls-Energie-Satz in der Elektrodynamik von Gustav Mie." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Nachrichten* (1914):23–36.

Born, Max, and Leopold Infeld. 1934. "Foundations of the New Field Theory." *Royal Society of London. Proceedings* A 144:425–451.

Caratheodory, Constantin. 1935. *Variationsrechnung und partielle Differentialgleichungen erster Ordnung*. Leipzig, Berlin: B. G. Teubner.

Corry, Leo. 1997. "David Hilbert and the Axiomatization of Physics (1894–1905)." *Archive for History of Exact Sciences* 51:83–198.

———. 1999a. "David Hilbert between Mechanical and Electromagnetic Reductionism (1910–1915)." *Archive for History of Exact Sciences* 53:489–527.

———. 1999b. "From Mie's Electromagnetic Theory of Matter to Hilbert's Unified Foundations of Physics." *Studies in History and Philosophy of Modern Physics* 30 B (2):159–183.

———. 1999c. "David Hilbert: Geometry and Physics (1900–1915)." In J. J. Gray (ed.), *The Symbolic Universe: Geometry and Physics (1890–1930)*, Oxford: Oxford University Press, 145–188.

———. 2004. *David Hilbert and the Axiomatization of Physics, 1898–1918: From "Grundlagen der Geometrie" to "Grundlagen der Physik"*. Dordrecht: Kluwer.

Corry, Leo, Jürgen Renn, and John Stachel (eds.). 1997. *Belated Decision in the Hilbert-Einstein Priority Dispute*. Vol. 278, Science.

CPAE 4: Martin J. Klein, A. J. Kox, Jürgen Renn, and Robert Schulmann (eds.). 1995. *The Collected Papers of Albert Einstein*. Vol. 4. *The Swiss Years: Writings, 1912–1914*. Princeton: Princeton University Press.

CPAE 6: A. J. Kox, Martin J. Klein, and Robert Schulmann (eds.). 1996. *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. Princeton: Princeton University Press.

CPAE 6E: *The Collected Papers of Albert Einstein*. Vol. 6. *The Berlin Years: Writings, 1914–1917*. English edition translated by Alfred Engel, consultant Engelbert Schucking. Princeton: Princeton University Press, 1996.

CPAE 8: Robert Schulmann, A. J. Kox, Michel Janssen, and József Illy (eds.). 1998. *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. Princeton: Princeton University Press.

CPAE 8E: *The Collected Papers of Albert Einstein*. Vol. 8. *The Berlin Years: Correspondence, 1914–1918*. English edition translated by Ann M. Hentschel, consultant Klaus Hentschel. Princeton: Princeton University Press, 1998.

Csikszentmihalyi, Mihaly. 1988. "Society, Culture and Person: a Systems View of Creativity." In R. J. Sternberg (ed.), *The Nature of Creativity*. Cambridge: Cambridge University Press.

Earman, John, and Clark Glymour. 1978. "Einstein and Hilbert: Two Months in the History of General Relativity." *Archive for History of Exact Sciences* 19:291–308.

Eddington, Arthur Stanley. 1923. *The Mathematical Theory of Relativity*. Cambridge: The University Press.

———. 1924. *The Mathematical Theory of Relativity*. 2nd ed. Cambridge: The University Press.

———. 1925. *Relativitätstheorie in Mathematischer Behandlung*. Translated by Alexander Ostrowski Harry Schmidt. Berlin: Springer.

Einstein, Albert. 1913. "Zum gegenwärtigen Stande des Gravitationsproblems." *Physikalische Zeitschrift* 14 (25):1249–1262. (English translation in volume 3 of this series.)

———. 1914a. "Prinzipielles zur verallgemeinerten Relativitätstheorie und Gravitationstheorie." *Physikalische Zeitschrift* 15:176–180.

———. 1914b. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1914) (XLI):1030–1085.

———. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLIV):778–786.

———.. 1915b. "Zur allgemeinen Relativitätstheorie (Nachtrag)." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVI):799–801.

———. 1915c. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVII):831–839.

———. 1915d. [Zusammenfassung der Mitteilung "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie."] *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVII):803.

———. 1915e. "Die Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915) (XLVIII–XLIX):844–847.

———. 1916a. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49 (7):769–822.

———. 1916b. "Hamiltonsches Prinzip und allgemeine Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1916) (XLII):1111–1116.

———. 1916c. "Näherungsweise Integration der Feldgleichungen der Gravitation." *Sitzung der physikalisch-mathematischen Klasse* 668–96. (CPAE 6, Doc. 32, 348–57)

Einstein, Albert, and Marcel Grossmann. 1914. "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie." *Zeitschrift für Mathematik und Physik* 63 (1 / 2):215–225.

Fölsing, Albrecht. 1997. *Albert Einstein: a biography*. New York: Viking.

Frei, Günther, ed. 1985. *Der Briefwechsel David Hilbert-Felix Klein (1886–1918)*. Vol. 19, *Arbeiten aus der Niedersächsischen Staats- und Universitätsbibliothek Göttingen*. Göttingen: Vandenhoeck & Ruprecht.

Goenner, Hubert. 2004. "On the History of Unified Field Theories." *Living Reviews of Relativity* 7 <http://www.livingreviews.org>.

Goenner, Hubert, Jürgen Renn, Jim Ritter, and Tilman Sauer (eds.). 1999. *The Expanding Worlds of General Relativity*. (*Einstein Studies* vol. 7.) Boston: Birkhäuser.

Guth, E. 1970. "Contribution to the History of Einstein's Geometry as a Branch of Physics." In *Relativity,* edited by M. Carmeli et al. New York, London: Plenum Press, 161–207.

Havas, Peter. 1989. "The Early History of the 'Problem of Motion' in General Relativity." In *Einstein and the History of General Relativity*, edited by Don Howard and John Stachel. (*Einstein Studies* vol. 1.) Boston: Birkhäuser, 234–276.

Hilbert, David. 1905. "Logische Prinzipien des mathematischen Denkens." Ms. Vorlesung SS 1905, annotated by E. Hellinger, Bibliothek des Mathematischen Seminars, Göttingen.

———. 1912–13. "Molekulartheorie der Materie." Ms. Vorlesung WS 1912–13, annotated by M. Born, Nachlass Max Born #1817, Stadtbibliothek Berlin.

———. 1913. "Elektronentheorie." Ms. Vorlesung SS 1913, Bibliothek des Mathematischen Seminars, Göttingen.

———. 1916. "Die Grundlagen der Physik. (Erste Mitteilung)." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1915):395–407. (English translation in this volume.)

———. 1917. "Die Grundlagen der Physik (Zweite Mitteilung)." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1917):53–76. (English translation in this volume.)

———. 1924. "Die Grundlagen der Physik." *Mathematische Annalen* 92:1–32.

———, ed. 1935. *Gesammelte Abhandlungen, Band III: Analysis, Grundlagen der Mathematik, Physik, Verschiedenes, Lebensgeschichte*. [1932–35, 3 vols.]. Berlin: Springer.

———. 1971. "Über meine Tätigkeit in Göttingen." In *Hilbert-Gedenkenband*, ed. K. Reidemeister. Berlin, Heidelberg, New York: Springer, 79–82.

Howard, Don, and John D. Norton. 1993. "Out of the Labyrinth? Einstein, Hertz, and the Göttingen Answer to the Hole Argument." In *The Attraction of Gravitation: New Studies in the History of General Relativity,* edited by John Earman, Michel Janssen and John D. Norton. Boston/Basel/Berlin: Birkhäuser, 30–62.

Janssen, Michel and Matthew Mecklenburg. 2006. "Electromagnetic Models of the Electron and the Transition from Classical to Relativistic Mechanics." In *Interactions: Mathematics, Physics and Philosophy, 1860–1930*, edited by V. F. Hendricks et al. *Boston Studies in the Philosophy of Science*, Vol. 251. Dordrecht: Springer, 65–134.

Kerschensteiner, Georg, ed. 1887. *Paul Gordan's Vorlesungen über Invariantentheorie. Zweiter Band: Binäre Formen*. Leipzig: Teubner.

Klein, Felix. 1917. "Zu Hilberts erster Note über die Grundlagen der Physik." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1917):469–482.

———. 1918a. "Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1918):171–189.

———. 1918b. "Über die Integralform der Erhaltungsätze und die Theorie der räumlich-geschlossenen Welt." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1918):394–423.

———. 1921. "Zu Hilberts erster Note über die Grundlagen der Physik." In *Gesammelte Mathematische Abhandlungen,* edited by R. Fricke and A. Ostrowski. Berlin: Julius Springer, 553–567.

Komar, Arthur. 1958. "Construction of a Complete Set of Independent Observables in the General Theory of Relativity." *Physical Review* 111:1182–1187.

Kretschmann, Erich. 1917. "Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie." *Annalen der Physik* 53 (16):575–614.

Laue, Max. 1911a. "Zur Dynamik der Relativitätstheorie." *Annalen der Physik* 35: 524–542.

———. 1911b. *Das Relativitätsprinzip*. Braunschweig: Friedrich Vieweg und Sohn.

Laue, Max von. 1920. "Theoretisches über neuere optische Beobachtungen zur Relativitätstheorie." *Physikalische Zeitschrift* 21:659–662.

Lichnerowicz, André. 1946. "Sur le caractère euclidien d'espaces-temps extérieurs statiques partout réguliers." *Academie des Sciences* (Paris). *Comptes Rendus* 222:432–436.

Lorentz, Hendrik A., et al. 1923. *The Principle of Relativity.* London: Methuen & Co.

Majer, Ulrich and Tilman Sauer. 2005. "'Hilbert's World Equations' and His Vision of a Unified Science." In *The Universe of General Relativity*, edited by A. Kox and J. Eisenstaedt. (*Einstein Studies*, vol. 11.) Boston: Birkhäuser, 259–276.

Mehra, Jagdish. 1974. *Einstein, Hilbert, and the Theory of Gravitation. Historical Origins of General Relativity Theory.* Dordrecht, Boston: D. Reidel Publishing Company.

Mie, Gustav. 1912a. "Grundlagen einer Theorie der Materie. Erste Mitteilung." *Annalen der Physik* 37:511–534. (English translation of excerpts in this volume.)

———. 1912b. "Grundlagen einer Theorie der Materie. Zweite Mitteilung." *Annalen der Physik* 39:1–40.

———. 1913. "Grundlagen einer Theorie der Materie. Dritte Mitteilung." *Annalen der Physik* 40:1–66. (English translation of excerpts in this volume.)

Noether, Emmy. 1918. "Invariante Variationsprobleme." *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten* (1918):235–257.

Norton, John D. 1984. "How Einstein Found His Field Equations, 1912–1915." *Historical Studies in the Physical Sciences* 14:253–316.

Pais, Abraham. 1982. *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*. Oxford, New York, Toronto, Melbourne: Oxford University Press.

Papapetrou, Achille. 1974. *Lectures on General Relativity.* Dordrecht/Boston: D. Reidel.

Pauli, Wolfgang. 1921. "Relativitätstheorie." In *Encyklopädie der mathematischen Wissenschaften, mit Einschluss ihrer Anwendungen*, edited by Arnold Sommerfeld. Leipzig: B. G. Teubner, 539–775.

———. 1958. *Theory of Relativity.* Translated by G. Field. London: Pergamon.

———. 1979. *Scientific Correspondence with Bohr, Einstein, Heisenberg, a.o. Volume 1: 1919–1929*. New York: Springer.

Reidemeister, Kurt, ed. 1971. *Hilbert-Gedenkenband*. Berlin, Heidelberg, New York: Springer,

Renn, Jürgen. 1994. "The Third Way to General Relativity." Preprint n° 9, *Max Planck Institute for the History of Science,* Berlin (http://www.mpiwg-berlin.mpg.de/Preprints/P9.PDF). (Revised edition in vol. 3 of this series.)

Renn, Jürgen, and Tilman Sauer. 1996. "Einsteins Züricher Notizbuch." *Physikalische Blätter* 52:865–872.

———. 1999. "Heuristics and Mathematical Representation in Einstein's Search for a Gravitational Field Equation." In (Goenner et al. 1999, 87–125).

Rosenfeld, Leon. 1940. "Sur le tenseur d'impulsion-énergie." *Mémoires de l'Academie royale de Belgique* 18 (16):1–30.

Rowe, David. 1989. "Klein, Hilbert, and the Göttingen Mathematical Tradition." *Osiris* 5:186–213.

———. 1999. "The Göttingen Response to General Relativity and Emmy Noether's Theorems." In *The Visual World: Geometry and Physics (1890–1930)*, ed., J. J. Gray. Oxford: Oxford University Press.

Sauer, Tilman. 1999. "The Relativity of Discovery: Hilbert's First Note on the Foundations of Physics." *Archive for History of Exact Sciences* 53:529–575.

———. 2002. "Hopes and Disappointments in Hilbert's Axiomatic 'Foundations of Physics.'" In *History of Philosophy and Science: new trends and perspectives*, ed. M. Heidelberger and F. Stadler. Dordrecht: Kluwer, 225–237.

Schaffner, Kenneth K. 1972. *Nineteenth-Century Aether Theories*. Oxford: Pergamon Press.

Schouten, Jan A. 1924. *Der Ricci-Kalkül*, 1st ed. Berlin: Springer-Verlag.

Schouten, Jan A., and Dirk J. Struik. 1935. *Algebra und Übertragungslehre*. Vol. 1, *Einführung in die neueren Methoden der Differentialgeometrie*. Groningen, Batavia: P. Noordhoff.

Schwarzschild, Karl. 1916. "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1916) (VII):189–196.

Siegmund-Schultze, Reinhard. 1998. *Mathematiker auf der Flucht vor Hitler: Quellen und Studien zur Emigration einer Wissenschaft*. Vol. 10, *Dokumente zur Geschichte der Mathematik*. Braunschweig Wiesbaden: Vieweg.

Slebodzinski, Wladyslaw. 1931. "Sur les equations de Hamilton." *Bulletin de l'Academie royale de Belgique* (5) (17):864–870.

Stachel, John. 1989. "Einstein's Search for General Covariance, 1912–1915." In *Einstein and the History of General Relativity,* edited by Don Howard and John Stachel. Boston/Basel/Berlin: Birkhäuser, 63–100.

———. 1992. "The Cauchy Problem in General Relativity - The Early Years." In *Studies in the History of General Relativity,* edited by Jean Eisenstaedt and A. J. Kox. Boston/Basel/Berlin: Birkhäuser, 407–418.

———. 1994. "Scientific Discoveries as Historical Artifacts." In *Current Trends in the Historiography of Science,* edited by Kostas Gavroglu. Dordrecht, Boston: Reidel, 139–148.

———. 1999. "New Light on the Einstein-Hilbert Priority Question." *Journal of Astrophysics and Astronomy* 20:91–101. Reprinted in (Stachel 2002).

———. 2002. *Einstein from 'B' to 'Z'*. (*Einstein Studies* vol. 9.) Boston: Birkhäuser.

Thorne, Kip S. 1994. *Black Holes and Time Warps: Einstein's Outrageous Legacy*. New York, London: Norton.

Vizgin, Vladimir P. 1989. "Einstein, Hilbert, and Weyl: The Genesis of the Geometrical Unified Field Theory Program." In *Einstein and the History of General Relativity,* edited by Don Howard and John Stachel. Boston/Basel/Berlin: Birkhäuser, 300–314.

———. 1994. *Unified Field Theories in the First Third of the 20th Century*. Translated by Barbour, Julian B. Edited by E. Hiebert and H. Wussing. Vol. 13, *Science Networks, Historical Studies*. Basel, Boston, Berlin: Birkhäuser.

Walter, Scott. 1999. "Minkowski, Mathematicians, and the Mathematical Theory of Relativity." In (Goenner et al. 1999, 45–86).

Weitzenböck, Roland. 1920. "Über die Wirkungsfunktion in der Weyl'schen Physik." *Akademie der Wissenschaften (Vienna). Mathematisch-naturwissenschaftliche Klasse. Sitzungsberichte* 129:683–696.

Weyl, Hermann. 1917. "Zur Gravitationstheorie." *Annalen der Physik* 54:117–145.

———. 1918a. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 1st ed. Berlin: Julius Springer.

———. 1918b. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 2nd ed. Berlin: Julius Springer.

———. 1918c. "Gravitation und Elektrizität." *Königlich Preußische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1918):465–480.

———. 1919. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 3rd, revised ed. Berlin: Julius Springer.

———. 1921. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 4th ed. Berlin: Julius Springer.

———. 1923. *Raum-Zeit-Materie. Vorlesungen über allgemeine Relativitätstheorie*. 5th, revised ed. Berlin: Julius Springer.

Whittaker, Edmund Taylor. 1951. *A History of the Theories of Aether and Electricity*. Vol. 1: *The Classical Theories*. London: Nelson.

TILMAN SAUER

EINSTEIN EQUATIONS AND HILBERT ACTION: WHAT IS MISSING ON PAGE 8 OF THE PROOFS FOR HILBERT'S FIRST COMMUNICATION ON THE FOUNDATIONS OF PHYSICS?[1]

1. INTRODUCTION

In contrast to Einstein's discovery of special relativity in 1905, his path towards the theory of general relativity is documented by a rich historical record. Not only did Einstein publish quite a few papers on earlier versions of a generalized theory of relativity, we also have a number of research manuscripts from crucial periods of his search, and we have an extensive correspondence from the relevant years. Hilbert's involvement in the discovery of general relativity is less abundantly documented but also here we have a few key documents that shed light on his work. Compared to other episodes in the history of science, the history of general relativity is very well written, and specifically the competition between Einstein and Hilbert in the final weeks before the publication of generally covariant field equations of gravitation in late 1915 has been commented on extensively.[2] Nevertheless, much of the historical literature on the Einstein-Hilbert competition took sides in what was perceived as a priority debate and it still seems worthwhile to come to a succinct and balanced assessment of the respective contributions of both authors in the final establishment of the general theory of relativity. In this respect, a set of proofs of Hilbert's relevant paper are of some significance and with those proofs the fact that a piece of them is missing. Although the fact that a piece of those proofs is missing is well known and was briefly commented on by several authors, the question naturally arises as to whether that missing part could have contained information that would compel us to reassess the historical account?

---

1 This paper was first published in *Archive for History of Exact Sciences* 59 (2005) 577–590, and is reprinted here with their kind permission.

2 See (Corry 2004; Corry, Renn, and Stachel 1997; Earman and Glymour 1978; Logunov, Mestvirishvili and Petrov 2004; Mehra 1974; Norton 1984; Pais 1982; Rowe 2001; Sauer 1999; Stachel 1999; Vizgin 2001), and "Hilbert's Foundation of Physics …" (in this volume), as well as further references cited in these works.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

2. THE CONTEXT

Before focussing on some minor yet significant details of the historical record, let me briefly review the broader historical context. In 1907, Einstein first formulated his equivalence hypothesis according to which no physical experiment can distinguish between the existence of a homogeneous, static gravitational field in a Newtonian inertial frame of reference and a uniformly and rectilinearly accelerated frame of reference that is free of any gravitational field. The hypothesis linked the problem of generalizing the special theory of relativity to accelerated motion with the problem of a relativistic theory of gravitation. In 1912, Einstein realized that such a relativistic theory of gravitation could not be achieved using a scalar gravitational potential but required the introduction of the metric tensor as the crucial mathematical object for a generalized theory of relativity. Together with his mathematician friend Marcel Grossmann, Einstein published an "Outline of a Generalized Theory of Relativity and a Theory of Gravitation" in 1913 (Einstein and Grossmann 1913). The theory of this "Outline" (*Entwurf*) has already many features of the final theory of general relativity except for one "dark spot." Einstein and Grossmann did not succeed in finding gravitational field equations for the components of the metric tensor that were both generally covariant and acceptable from the point of view of Einstein's understanding of the requirements for a satisfactory theory of gravitation.

The final episode of Einstein's path towards General Relativity began in the fall of 1915 when Einstein lost faith in the validity of the field equations of his "Outline" and reverts to a reassessment of the mathematics of general covariance as developed in the work of Riemann, Christoffel, Ricci and Levi-Civita. The final steps were taken in four successive communications to the Prussian Academy of Sciences, all of them presented for publication in the month of November 1915 (Einstein 1915a, b, c, d). On November 4, Einstein advanced field equations that are based on the Ricci tensor but that are not yet generally covariant (Einstein 1915a). Instead, by stipulation of a restrictive condition on the admissible coordinates, he split off a part of the Ricci tensor and equated the remaining part to an unspecified energy-momentum tensor as the source of the gravitational field. In an addendum to this paper, presented a week later on November 11 (Einstein 1915b), Einstein temporarily entertains the speculation that all matter might be of electromagnetic origin. This assumption allowed him to advance a generally covariant field equation of gravitation where the Ricci tensor is directly set proportional to the energy-momentum tensor. Another week later, Einstein presented a paper to the Berlin Academy in which he successfully computed the anomalous advance of the perihelion of Mercury on the basis of his new equations (Einstein 1915c). And yet another week later, Einstein realized that he can add a trace term to the right-hand side of his field equations which turns them into what we now refer to as the Einstein equations (Einstein 1915d).

David Hilbert's path towards general relativity is a rather different one. Half a generation older than Einstein, Hilbert in 1900 formulated his famous 23 problems of mathematical research of the coming century to the International Congress of Mathematicians in Paris. The sixth of these problems asked for an axiomatization of phys-

ics. After working on the theory of integral equations in the first decade of the century, Hilbert himself then turned to an intense study of all fields of theoretical physics. In the course of his study of contemporary physics literature he soon became interested in an attempt by the German physicist Gustav Mie to generalize Maxwellian electrodynamics so as to turn it into a theory of matter. Mie's idea was to take Maxwellian electrodynamics in its variational formulation but to search for a generalized Lagrangian entering the action, keeping the requirement of Lorentz covariance but allowing for the Lagrangian to depend explicitly on the electromagnetic vector potential. Mie's hope was to find a modified Lagrangian that would produce modified Maxwell equations which, on microscopic scales, would allow for particle-like solutions. Around that time, Hilbert also became interested in Einstein's recent work on a relativistic theory of gravitation and invited Einstein to give a series of lectures on his new theory to the Göttingen mathematicians and physicists. After Einstein presented his theory in Göttingen in July 1915, Hilbert left Göttingen for his summer vacations and began pondering on Einstein's "Outline" theory. Shortly after coming back to Göttingen at the beginning of the winter term, Hilbert himself then presented a paper to the Göttingen Academy of Sciences. In this communication, Hilbert presented a theory of the "Foundations of Physics" which combined Mie's idea of a generalized electrodynamics with Einstein's idea of a generally covariant theory of gravitation.

The dateline on Hilbert's *First Communication* on the *Foundations of Physics* (Hilbert 1915) says that it was presented to the Göttingen Academy of Sciences on 20 November 1915. The dateline on Einstein's note on *The Field Equations of Gravitation* (Einstein 1915d) says that it was presented to the Berlin Academy of Sciences on 25 November 1915. From a comparison of the two publications, it appears that Hilbert preceded Einstein with the publication of the final gravitational field equations of general relativity by five days, notwithstanding the fact that both authors arrived at these equations along very different routes.

The question as to where the correct field equations of gravitation are first found in print is in need of some qualification. The gravitational field equations of general relativity may be written in two very different yet essentially equivalent ways. Einstein published his final field equations of 25 November (Einstein 1915d, 845),

$$G_{im} = -\kappa\left(T_{im} - \frac{1}{2}g_{im}T\right), \tag{1}$$

as an explicit set of differential equations for the components of the metric tensor g_{im}. Using the Ricci tensor G_{im} as the differential operator acting on the metric and the energy-momentum tensor T_{im} in the source term on the right-hand side made sure that his equations retained its form under arbitrary coordinate transformations, i.e. made them generally covariant. Adding a trace term $-(1/2)g_{im}T$ where $T = \Sigma g^{\rho\sigma}T_{\rho\sigma}$ to the right-hand side of his equations in his last November paper did not violate this feature. Hilbert published the gravitational field equations in implicit form in terms of a variational principle. He axiomatically postulated an action integral (Hilbert 1915, 396)

$$\int H \sqrt{g} d\omega, \tag{2}$$

where $g = |g_{\mu\nu}|$, $d\omega = dw_1 dw_2 dw_3 dw_4$ for spacetime coordinates w_i and required that the Lagrangian H that enters into the action of his variational formulation be invariant under arbitrary coordinate transformations. He also assumed that the Lagrangian splits into the sum of two parts, a gravitational part given by the Riemann curvature scalar and a matter part which he left unspecified except for the postulation that it depend only on the components of the metric and the components of the electromagnetic vector potential and its first derivatives. This specification technically renders Einstein's equations equivalent to Hilbert's action, except for some ambiguity in the assumptions on how the source term is to be specified, i.e. on the fundamental constitution of matter. Both Hilbert and Einstein had left the matter term undetermined to some extent. Einstein had not specified his source term at all. Hilbert had axiomatically required that the source term depend only on the electromagnetic variables and hence that all matter is of electromagnetic origin.

But several years ago it was pointed out (Corry, Renn and Stachel 1997) that a set of proofs for Hilbert's *First Communication* is extant in the Hilbert archives in Göttingen. It bears a printer's stamp of December 6, 1915, and differs in some significant respects from the published version.[3] The main difference pertains to a different treatment of the energy concept that motivated an axiomatic restriction of the general covariance of Hilbert's theory and that was substantially rewritten for the published version. In the published paper, the discussion of the energy concept no longer results in the postulation of a restriction of the general covariance. It was also pointed out that the proofs did not contain the explicit version of the gravitational field equations in terms of the Einstein tensor as does Hilbert's published paper. What we now call the Einstein tensor is obtained by adding a trace term to the Ricci tensor, its covariant divergence vanishes identically, and it is obtained from the explicit variation of the gravitational part of Hilbert's action integral. To be precise, in Einstein's paper of 25 November the trace term was added on the right-hand side of the field equation to the source term and not to the Ricci tensor on the left hand side and strictly speaking his paper does not contain the Einstein tensor explicitly but this difference is a minor detail since both variants are trivially equivalent. In view of the differences between the proofs and the published paper general agreement seems to have been reached[4] about the conclusion that the proofs unequivocally rule out the possibility that Einstein may have taken the clue of adding a trace term to his field equations of 11 November (Einstein 1915b) from Hilbert's paper (1915). No agreement, however, was reached on the question as to the path along which Hilbert arrived at his finally

3 Hilbert's paper was eventually issued only on March 31, 1916, but off-prints of the final version were available to Hilbert already by mid-February (Sauer 1999, note 74). Einstein's November papers were each published a week after their presentation to the Prussian Academy.

4 See (Corry 2004; Rowe 2001; Sauer 1999; Stachel 1999; Vizgin 2001) and also "Hilbert's Foundation of Physics …" (in this volume).

published theory: by taking the main clues from Einstein's paper, as suggested in (Corry, Renn and Stachel 1997), or along an independent logic of discovery, as first advocated in explicit response to this claim in (Sauer 1999). It also remains an open question to what extent Einstein in those weeks of October and November 1915 had heard directly or indirectly about Hilbert's work on his theory and to what extent he may have been influenced by what he heard, e.g. in entertaining temporarily the speculation that all matter is of electromagnetic origin.

To add to the complexity of the issue, it so happens that a portion of one sheet of the extant proofs for Hilbert's *First Communication* is missing.[5] In view of this fact, it seems worthwhile to discuss the question what part of the argument of the proofs is missing and whether an answer to this question may possibly affect our assessment of the Einstein-Hilbert competition in late 1915. In the following, I will argue that an analysis of the internal structure of the text and argument of the proofs and the published version of Hilbert's paper shows that the missing piece in all probability did not contain an explicit version of the Einstein tensor and its trace term. The analysis rather suggests that it contained an explicit form of the Riemann curvature scalar and the Ricci tensor as a specification of the Lagrangian in Hilbert's variational principle.

3. WHAT IS MISSING IN THE PROOFS

Axiom I of Hilbert's *First Communication*, as presented on page 2 of his proofs,[6] introduces an action integral[7]

$$\int H\sqrt{g}\,d\tau \tag{3}$$

where $g = |g_{\mu\nu}|$, $d\tau = dw_1 dw_2 dw_3 dw_4$ and H is a Lagrangian density that depends on the components of the metric $g_{\mu\nu}$, its first and second derivatives with respect to the coordinates w_1 of the spacetime manifold, $g_{\mu\nu l} = \partial g_{\mu\nu}/\partial w_l$ and $g_{\mu\nu lk} = \partial^2 g_{\mu\nu}/\partial w_l \partial w_k$, respectively, and also depends on the components of the electromagnetic vector potential q_s and its first derivatives $q_{sl} = \partial q_s/\partial w_l$. Specifically, the axiom demands that the laws of physics be given by the vanishing of the

5 See (Sauer 1999, note 75) and "Hilbert's Foundation of Physics …" note 6 (in this volume).

6 Niedersächsische Staats- und Universitätsbibliothek (NSUB), Handschriftenabteilung, Cod. Ms. Hilbert 634, f.23-29. Facsimile versions of both Hilbert's proofs and of the published version were made available online by the Max Planck Institute for the History of Science, Berlin, on <http://echo.mpiwg-berlin.mpg.de/content/relativityrevolution/hilbert>. A facsimile of the published version is also available online from the website of the *Göttinger Digitalisierungszentrum* of the NSUB, see <http://gdz.sub.uni-goettingen.de/gdz>.

7 The argument being partly one of textual exegesis, I am keeping strictly to Hilbert's notation. He uses an imaginary time-coordinate and, following standard usage of the time, refers to the Lagrangian density as a Hamiltonian function. Contrary to later and current usage, Hilbert and Einstein at the time also consistently wrote contravariant indices of coordinate differentials as subscript indices. Hilbert also uses subscript indices to denote partial coordinate derivatives without, however, indicating this meaning by separating the index with a comma.

variation of the action integral with respect to the fourteen potentials $g_{\mu\nu}$ and q_s for some as yet unspecified function H.

Axiom II, immediately following, then demands that H must be an invariant under all coordinate transformations. Other than that, the Lagrangian H is left undetermined by the axioms.

On page 3, Hilbert writes down the "ten Lagrangian differential equations"[8]

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_k} - \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}}, \qquad (\mu,\nu = 1,2,3,4) \qquad \text{(4-pr)}$$

which he calls the "fundamental equations of gravitation," and the four Lagrangian differential equations

$$\frac{\partial\sqrt{g}H}{\partial q_h} = \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}}, \qquad (h = 1,2,3,4) \qquad \text{(5-pr)}$$

which he calls the "fundamental equations of electrodynamics or the generalized Maxwell equations." Hilbert then proceeds to discuss the concept of energy in the theory by looking at what we would now call Lie variations of the action, i.e. variations of the metric that arise from pure coordinate transformations. In the course of this discussion he introduces the notational "abbreviation"

$$[\sqrt{g}H]_{\mu\nu} = \frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_k} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}} \qquad (4)$$

which he calls "the Lagrangian variational derivative of $\sqrt{g}H$ with respect to $g^{\mu\nu}$." He observes that the fundamental equations of gravitation (4-pr) may now compactly be written as

$$[\sqrt{g}H]_{\mu\nu} = 0. \qquad \text{(8-pr)}$$

Hilbert's discussion of the energy concept in the proofs does not provide any further specifications of the Lagrangian H, although it does lead to a third axiom that restricts the covariance of the generally covariant equations (4-pr), (5-pr), by demanding that the physically admissible coordinates for the theory obey a set of equations that are not generally covariant.[9]

It is towards the end of the discussion of the problem of the energy concept and the significance of his third axiom, which runs until the bottom of page 7, that we find two passages missing in the proofs, since the top portion of the sheet that contains pages 7and 8 was cut off.[10] Without any further discussion of Hilbert's treatment of

8 Hilbert tended to use equation numbers only for those equations that he actually referred to in his text. I will use his own equation numbers whenever an equation was given one and indicate this fact by adding "-pr" resp. "-pu" to the number, depending on whether it is the equation number used in the proofs or the published version, respectively.

the energy concept,[11] I will assume that the missing portion on the top of page 7, i.e. on the verso of the top of page 8, is not in any way relevant to the question under investigation in this note. But what is missing on page 8?

On page 8 of the proofs, immediately following the excised portion, Hilbert asserts: "Since K depends only on $g^{\mu\nu}$, $g^{\mu\nu}{}_k$, $g^{\mu\nu}_{lk}$, the ansatz (17-pr) allows us to express the energy E [...] solely as a function of the gravitational potentials $g^{\mu\nu}$ and their derivatives, if only we assume L not to depend on $g^{\mu\nu}_s$, but only on $g^{\mu\nu}$, q_s, q_{sk}." In the next sentence, Hilbert states that he would make that latter assumption in the following.

We observe that the quantities K and L had not been used earlier in the proofs,[12] and we may conclude that K must have been introduced just before as a function of the components of the metric and its derivatives only, and that L must have been introduced just before as a function of the electromagnetic potential, its derivatives as well as of the components of the metric and its first derivatives, although the dependence of L on the derivatives of the metric is immediately assumed away for the rest of the text. We also observe that the previous page has an equation that is numbered (16-pr) and that the next line gives an equation that is numbered (18-pr). The equation with number (17-pr) is referred to a few pages later, on page 11, where Hilbert writes that "because of (17-pr)" the fundamental equations of gravitation (8-pr) take the form

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial \sqrt{g}L}{\partial g^{\mu\nu}} = 0, \tag{26-pr}$$

and the fundamental equations of electrodynamics take the form

$$[\sqrt{g}L]_h = 0. \tag{27-pr}$$

Spelling out $[\sqrt{g}K]_{\mu\nu}$ in terms of the definition (4), eq. (26-pr) reads

---

9 Contrary to the discussion in (Logunov, Mestvirishvili and Petrov 2004), this condition is conceptually very different from what we now call a coordinate condition since it pertains to *any* possible application of the field equations. In these volumes, such restricting equations are called "coordinate restrictions" as opposed to "coordinate conditions." Nonetheless, there is a significant difference between Einstein's use of "coordinate restrictions" prior to his final version of the general theory of relativity and Hilbert's third axiom in the proofs. Einstein used "coordinate restrictions" to derive field equations that are covariant only under a correspondingly restricted group of coordinate transformations. Hilbert kept the generally covariant field equations as fundamental field equations and only postulated a limitation of the physically admissible coordinate systems.

10 For a description of the physical appearance of the proofs, see (Sauer 1999, note 75).

11 See (Sauer 1999) and "Hilbert's Foundation of Physics …" (in this volume).

12 The choice of characters seems to have been motivated by alphabetical order. After denoting the generic "Hamiltonian function" as H, some invariant expression is denoted on page 4 as $J^{(h)}$. Later, on page 10, the electromagnetic field tensor is denoted by $M_{ks} = q_{sk} - q_{ks}$.

$$\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}} + \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}_k} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}_{kl}} = 0. \tag{5}$$

Assuming that the missing piece introduced the quantities K and L by specifying H as some function of these quantities, $H = H(K, L)$, and taking into account that $L = L(g^{\mu\nu}, q_s, q_{sk})$ was assumed not to depend on $g^{\mu\nu}_k$ and $g^{\mu\nu}_{kl}$ we conclude that, in all probability, eq. (17-pr) must have been of the form:

$$H = \zeta(K + L) \tag{6}$$

with some constant ζ that may well have been set equal to 1. Clearly, Eq. (27-pr) is consistent with this conclusion. We also note that later in the text the quantities K and L are referred to as "invariants" (L on page 9 and on page 10, K on page 11).

Taking together these bits of information from the text of the proofs, we can draw the following preliminary conclusions about the content of the missing piece:

1. It must have contained an equation of the form (6) that was given the number (17-pr).
2. The missing piece introduced a quantity K in such a way that the definition or characterization of K, whatever it was, implied that $K = K(g^{\mu\nu}, g^{\mu\nu}_l, g^{\mu\nu}_{kl})$ is an invariant and only depends on the components of the metric and its first and second derivatives.
3. The missing piece introduced a quantity L in such a way that the definition or characterization of L, whatever it was, implied that $L = L(q_s, q_{sl}, g^{\mu\nu}, g^{\mu\nu}_k)$ is an invariant and depends on the components of the electromagnetic vector potential and its first derivatives as well as on the metric components and its first derivatives.

It should be noted that these conclusions emerge from looking at the existing text of the proofs alone, without taking recourse to the published version or any other historical source.

4. WHAT IS CONTAINED IN THE PUBLISHED VERSION

Let us now take further account of Hilbert's published version of his *First Communication* (Hilbert 1915). As was indicated above, the published version differs significantly from the proofs in several respects, the main difference being a completely revised discussion of the energy theorem. Specifically, with respect to the gravitational and electrodynamical field equations, however, the differences are not significant, as we will see, apart from the fact that the explicit evaluation of the variational derivative of the gravitational part of the Lagrangian K is found only in the published version and not in the existing part of the proofs. Whether it may have been on the missing part of the proofs will be discussed below.

The formulation of the first two axioms is the same, and in the published version, Hilbert again wrote down the fundamental equations (4-pr), and (5-pr), albeit in a slightly different form as

$$\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_k} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}H}{\partial g^{\mu\nu}_{kl}} = 0, \tag{4-pu}$$

and

$$\frac{\partial\sqrt{g}H}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}H}{\partial q_{hk}} = 0. \tag{5-pu}$$

The equivalence of eqs. (4-pr) and (5-pr), with (4-pu) and (5-pu) is, of course, completely trivial but the form (4-pu), (5-pu) allowed Hilbert to introduce the abbreviated notation $[\sqrt{g}H]_{\mu\nu}$ and $[\sqrt{g}H]_h$ already at this point as the left hand sides of the "fundamental equations" (4-pu) and (5-pu).

The specification of the Lagrangian H in terms of a gravitational part K and an electromagnetic part L appears twice in the published version. The first time the relevant equation appears it is in a context that would fit quite naturally into the missing piece of page 8 of the proofs. The relevant passage reads:

> As far as the world function H is concerned, further axioms are needed to determine its choice in a unique way. If the gravitational field equations are to contain only second derivatives of the potentials $g^{\mu\nu}$, then H must have the form
>
> $$H = K + L \tag{7}$$
>
> where K is the invariant that derives from the Riemannian tensor (curvature of the four-dimensional manifold)
>
> $$K = \sum_{\mu\nu} g^{\mu\nu} K_{\mu\nu} \tag{8}$$
>
> $$K_{\mu\nu} = \sum_\kappa \left(\frac{\partial}{\partial w_\nu}\left\{ \begin{matrix} \mu\kappa \\ \kappa \end{matrix} \right\} - \frac{\partial}{\partial w_\kappa}\left\{ \begin{matrix} \mu\nu \\ \kappa \end{matrix} \right\} \right) + \sum_{k,\lambda}\left(\left\{ \begin{matrix} \mu\kappa \\ \lambda \end{matrix} \right\}\left\{ \begin{matrix} \lambda\nu \\ \kappa \end{matrix} \right\} - \left\{ \begin{matrix} \mu\nu \\ \lambda \end{matrix} \right\}\left\{ \begin{matrix} \lambda\kappa \\ \kappa \end{matrix} \right\} \right) \tag{9}$$
>
> and where L only depends on $g^{\mu\nu}, g^{\mu\nu}_l, q_s, q_{sk}$. (Hilbert 1915, 402)

Hilbert then adds the following sentence: "Finally, we will, in the following, make the simplifying assumption that L does not depend on $g^{\mu\nu}_l$." The physical size of the missing piece allows for some ten lines of text or the equivalent of some smaller number of lines of text plus a number of displayed equations, taking into account that a displayed equation would take up more than a single line of text.[13] In view of this restriction, the passage in the published version is clearly too long to be inserted into

13 See (Sauer 1999, note 75), the length of the type area seems to vary slightly over the different pages of the proofs.

the missing piece of the proofs. However, we can easily cut down the passage to fit into the size of the missing piece as, e.g., with the following German sentence:

> Wir machen im folgenden den Ansatz
>
> $$H = K + L \tag{10}$$
>
> wo K die aus dem Riemannschen Tensor entspringende Invariante
>
> $$K = \sum_{\mu\nu} g^{\mu\nu} K_{\mu\nu} \tag{11}$$
>
> $$K_{\mu\nu} = \sum_{\kappa} \left(\frac{\partial}{\partial w_\nu} \left\{ \begin{matrix} \mu\kappa \\ \kappa \end{matrix} \right\} - \frac{\partial}{\partial w_\kappa} \left\{ \begin{matrix} \mu\nu \\ \kappa \end{matrix} \right\} \right) + \sum_{k,\lambda} \left(\left\{ \begin{matrix} \mu\kappa \\ \lambda \end{matrix} \right\} \left\{ \begin{matrix} \lambda\nu \\ \kappa \end{matrix} \right\} - \left\{ \begin{matrix} \mu\nu \\ \lambda \end{matrix} \right\} \left\{ \begin{matrix} \lambda\kappa \\ \kappa \end{matrix} \right\} \right) \tag{12}$$
>
> bedeutet und L nur von $g^{\mu\nu}$, $g^{\mu\nu}_l$, q_s, q_{sk} abhängt.[14]

It seems perfectly natural to assume that this passage or some very similar variant of it was the missing piece on page 8 of the proofs. And, as already conjectured in (Sauer 1999, note 82), Hilbert himself may have cut out this piece from his proofs, perhaps to paste it into some other unknown manuscript of his, e.g. into the manuscript for his revised version.

As indicated above, the equation $H = K + L$ appears at one other place in the published version of Hilbert's *First Communication*. This passage reads:

> It remains to show directly how with the assumption
>
> $$H = K + L \tag{20-pu}$$
>
> the generalized Maxwell equations (5-pu) put forth above are entailed by the gravitational equations (4-pu).
>
> Using the notation introduced earlier for the variational derivatives with respect to the $g^{\mu\nu}$, the gravitational equations, because of (20-pu), take the form
>
> $$[\sqrt{g}K]_{\mu\nu} + \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} = 0. \tag{21-pu}$$
>
> The first term on the left hand side becomes
>
> $$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2} K g_{\mu\nu} \right), \tag{13}$$
>
> as follows easily without calculation from the fact that $K_{\mu\nu}$, apart from $g_{\mu\nu}$, is the only tensor of second rank ("Ordnung") and K the only invariant, that can be formed using only the $g^{\mu\nu}$ and their first and second differential quotients $g^{\mu\nu}_k$, $g^{\mu\nu}_{kl}$. (Hilbert 1915, 404 f.)

14 "We now make the ansatz [...] where K is the invariant that derives from the Riemannian tensor [...] and where L only depends on $g^{\mu\nu}$, $g^{\mu\nu}_l$, q_s, q_{sk}."

And after this assertion, Hilbert adds the following comment as to the apparent equivalence of his equations to those published by Einstein:

> The resulting differential equations of gravitation are, it seems to me, in agreement with the broad ("großzügigen") theory of general relativity established by Einstein in his later papers.

The reference to Einstein's "later papers" is specified in a footnote by citing all four of Einstein's November memoirs (Einstein 1915a, b, c, d) including the last one that was presented to the Berlin Academy only on 25 November (Einstein 1915d). The question arises whether the missing piece of the proofs could have contained equation (13), i.e. the explicit form of the variational derivative for some gravitational Lagrangian K. Specifically under the assumption that K was defined or characterized as the Riemannian curvature scalar, it would then have displayed what we now call the Einstein tensor with its trace term $-\frac{1}{2}Kg_{\mu\nu}$. This reading would allow revival of a speculation that a version of the theory as laid out in the proofs may then possibly have inspired Einstein to make the transition of his field equations of his second November memoir of 11 November 1915 (Einstein 1915b) to those of his final November paper of 25 November 1915 (Einstein 1915d) by adding a similar trace term to the matter term of his previous equation.

However, from the internal logic and structure of both the argument in the proofs and in the published version, this conjecture seems highly unlikely for the following reasons. In addition to equation (13) or some similar equation displaying the explicit form of the variational derivative of the gravitational part of the Lagrangian, the missing piece must still have contained an equation of the form (6), as in (20-pu), and some kind of characterization of the quantities K and L as discussed above on the basis of the proofs alone. In addition, it must also have contained some kind of characterization of the term $K_{\mu\nu}$ which appears in equation (13) but which had not appeared in the proofs before. In view of the physical size of the missing piece, the explicit form of the Ricci tensor $K_{\mu\nu}$, as in (12), could hardly have fitted on it in addition to equation (20-pu), as well as equation (13). Therefore, the quantity K must then have been defined or characterized without using its explicit form, maybe only with words ("die aus dem Riemannschen Krümmungstensor $K_{\mu\nu}$ entspringende Invariante K").

However, there are at least two arguments against the assumption that the missing piece contained equation (13) in addition to equation (20-pu) and some minimal information needed to introduce K and L.

1. Nowhere in the extant parts of the proofs does Hilbert calculate explicitly the result of the variational derivative or argues on this level. Indeed, in and of itself such an explicit calculation would be at odds with the general thrust of his communication which is to draw quite general conclusions from combining variational calculus and invariant theory. And in the published version, the explicit form of the variational derivative of the gravitational part of the Lagrangian is clearly directly motivated by Hilbert's comment on the presumed equivalence of

his own equations with those of Einstein's November memoirs, specifically as it seems with the final ones of 25 November 1915.

2. The mathematical assertion captured by equation (13), i.e. the assertion that the Einstein tensor $K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}$ is obtained by a variation of the Riemann curvature scalar K with respect to the metric $g^{\mu\nu}$, must have been given with even less comments on how this result is obtained and on what assumptions are needed for its validity, as were given in the published version.

To elaborate on the second point, let me finally comment on the derivation of the Einstein tensor from a variation of the Riemann curvature. As pointed out in (Corry, Renn and Stachel 1997), the fact that Hilbert's assertion quoted above about the uniqueness of the Einstein tensor, if taken literally, is wrong, since there are many invariants that are of second rank and "can be formed using only the $g_{\mu\nu}$ and their first and second differential quotients." However, earlier on, Hilbert had also mentioned the condition that second derivatives are to be contained in the gravitational equations only linearly. This additional condition fixes the tensor to the form $K_{\mu\nu} - \alpha K g_{\mu\nu}$ with some undetermined factor α. This factor α is determined to be equal to 1/2 if it is further assumed that the covariant divergence of the expression vanishes, an assumption that is never mentioned explicitly in the published version, although it is implied by the contracted Bianchi identities that follow from Hilbert's proto-version of Noether's second theorem in his published communication (Sauer 1999, note 104 and p. 564; Logunov, Mestvirishvili and Petrov 2004). Corry, Renn and Stachel (1997) also point out that, while Hilbert asserts that the result follows "without calculation," he does give a more explicit derivation of the Einstein tensor in his 1924 republication of his *Communications* on the *Foundations of Physics* (Hilbert 1924).[15] Nevertheless, we have contemporary evidence that may give a meaning to Hilbert's assertion. It is found in a letter by the mathematician Hermann Vermeil to Felix Klein, dated 2 February 1918.[16] In it Vermeil explicitly addressed the question how the result can be obtained "without calculation." The answer that he found goes like this:

Assuming that

$$[\sqrt{g}K]_{\mu\nu} \propto \sqrt{g}(K_{\mu\nu} - \alpha K g_{\mu\nu}) \tag{14}$$

15 I disagree with the claim in (Corry, Renn and Stachel 1997) that the 1924 republication was primarily motivated by Hilbert's wish to correct some errors of his 1915 publication. As argued elsewhere (Majer and Sauer 2005), it was on the contrary Hilbert's intention to reaffirm his own priority of the field equations after Einstein in his 1923 papers on Eddington's unified field theory had arrived at equations that were essentially equivalent to the gravitational field equations of 1915 in variational form in the context of the unified field theory program.

16 NSUB Cod. Ms. Klein 22B, f. 28. This letter was discussed extensively at a history of mathematics conference at Oberwolfach in May 2000 in which the Einstein-Hilbert competition was a central topic of discussion. The argument is also presented, apparently without knowledge of Vermeil's letter, in (Logunov, Mestvirishvili and Petrov 2004, 611). For Vermeil's role, see also the discussion in (Rowe 2001, 417f.).

which, as discussed, follows from Hilbert's assumptions if one also demands that second derivatives occur only linearly, Vermeil evaluated $[\sqrt{g}K]_{\mu\nu}$ (see (4) for the scalar $K = \sum_{\rho\sigma} g^{\rho\sigma} K_{\rho\sigma,}$ see (8), and obtained

$$[\sqrt{g}K]_{\mu\nu} = K\frac{\partial\sqrt{g}}{\partial g^{\mu\nu}} + \sqrt{g}K_{\mu\nu} + \sqrt{g}g^{\rho\sigma}\frac{\partial K_{\rho\sigma}}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}_k} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}_{kl}}. \tag{15}$$

Using $dg = -gg_{\mu\nu}dg^{\mu\nu}$, this turns into

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\right) + \sqrt{g}g^{\rho\sigma}\frac{\partial K_{\rho\sigma}}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}_k} + \sum_{k,l}\frac{\partial^2}{\partial w_k \partial w_l}\frac{\partial\sqrt{g}K}{\partial g^{\mu\nu}_{kl}}. \tag{16}$$

where all terms on the second line do not produce terms of the form (14).

While this derivation shows that Hilbert's claim in the published version about the derivation of the Einstein tensor is correct (granting that the postulate that second derivatives occur only linearly was implied) and credible, the question still remains as to why Hilbert should have done this derivation and included its result into the proofs without elaborating at all about the necessary steps and assumptions. Assuming that Hilbert added the explicit evaluation of $[\sqrt{g}K]_{\mu\nu}$ into the published version after seeing the explicit field equations of Einstein's final November paper, on the other hand, makes good sense. Let us not forget after all, that Hilbert in this context does cite Einstein's paper of 25 November.

5. CONCLUDING REMARKS

What was on the excised piece? Merely requiring continuity with the remaining text constrains the possibilities quite considerably. It is highly unlikely that the missing part contained the explicit result of a variational derivative of the action with respect to the metric and specifically some version of the Einstein tensor. Consistency with the remaining text rather leads virtually uniquely to the conclusion that on the missing piece Hilbert had specified the Lagrangian of his variational principle as a sum of a gravitational part and a matter part, that he had further specified the gravitational part as the Riemann curvature scalar, and that he did so by giving the Ricci tensor in its explicit form.

It still remains true that the proofs of Hilbert's *First Communication* on the *Foundations of Physics* already contain the correct gravitational field equations of general relativity in implicit form, i.e. in terms of a variational principle and the Hilbert action. The variational formulation is fully equivalent to the explicit Einstein equations published by Einstein a few days later, although the theory of Hilbert's proofs

was not yet a fully generally covariant theory. It remains an interesting task to spell out in detail a scenario by which Hilbert would have overcome the restriction implied by the third axiom of his proofs following his own heuristics and logic of discovery.

REFERENCES

Corry, Leo. 2004. *David Hilbert and the Axiomatization of Physics. From 'Grundlagen der Geometrie' to 'Grundlagen der Physik'*. Dordrecht/Boston/London: Kluwer.

Corry, Leo, Jürgen Renn and John Stachel. 1997. "Belated Decision in the Hilbert-Einstein Priority Dispute." *Science* 278: 1270–1273.

Earman, J., and C. Glymour. 1978. "Einstein and Hilbert. Two Months in the History of General Relativity." *Archive for History of Exact Sciences* 19: 291–308.

Einstein, Albert. 1915a. "Zur allgemeinen Relativitätstheorie." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 778–786, (CPAE 6, Doc. 21).

———. 1915b. "Zur allgemeinen Relativitätstheorie. (Nachtrag)." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 799–801, (CPAE 6, Doc. 22).

———. 1915c. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie." *Sitzungsberichte der Preußischen Akademie der Wissenschaften: 2. Halbband* XLVII: 831-839, (CPAE 6, Doc. 24).

———. 1915d. "Die Feldgleichungen der Gravitation." *Königlich Preußische Akademie der Wissenschaften* (Berlin). *Sitzungsberichte*: 844–847, (CPAE 6, Doc. 25).

Einstein, Albert, and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. Leipzig/Berlin: Teubner, (CPAE 4, Doc. 13).

Hilbert, David. 1915. "Die Grundlagen der Physik. (Erste Mitteilung.)" *Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. Nachrichten*, 395–407. (English translation in this volume.)

———. 1924. "Die Grundlagen der Physik." *Mathematische Annalen* 92: 1–32.

Logunov, A.A., M.A. Mestvirishvili und V.A. Petrov. 2004. "How Were the Hilbert-Einstein Equations Discovered?" *Physics-Uspekhi* 47 (2004) 607–621.

Majer, Ulrich and Sauer, Tilman. 2005. "Hilbert's World Equations and His Vision of a Unified Science." In A. J. Kox and J. Eisenstaedt (eds.) *The Universe of General Relativity*, 259–276.(*Einstein Studies*, vol. 11). Boston/Basel/Berlin: Birkhäuser.

Mehra, Jagdish. 1974. *Einstein, Hilbert, and The Theory of Gravitation*. Dordrecht/Boston: D. Reidel.

Norton, John. 1984. "How Einstein Found His Field Equations: 1912–1915." *Historical Studies in the Physical Sciences* 14 (1984) 253–316. Reprinted in D. Howard and J. Stachel (eds.) *Einstein and the History of General Relativity*. Boston: Birkhäuser, 101–159.

Pais, Abraham. 1982. *'Subtle is the Lord ...' The Science and the Life of Albert Einstein*. Oxford and New York: Clarendon Press and Oxford University Press.

Rowe, David. 2001. "Einstein Meets Hilbert: At the Crossroads of Physics and Mathematics." *Physics in Perspective* 3: 379–424.

Sauer, Tilman. 1999. "The Relativity of Discovery. Hilbert's First Note on the Foundations of Physics." *Archive for History of Exact Sciences* 53: 529–575.

Stachel, John. 1999. "New Light on the Einstein-Hilbert Priority Question." *Journal of Astrophysics and Astronomy* 20: 91-101. Reprinted in: Stachel, John. *Einstein from 'B' to 'Z'*. Boston/Basel/Berlin: Birkhäuser, 2002, 353–364.

Vizgin, V. P. 2001. "On the discovery of the gravitational field equations by Einstein and Hilbert: new materials." *Physics-Uspekhi* 44: 1283–1298.

First proof of my first note.

The Foundations of Physics.

(First communication.)[1]

by

David Hilbert.

Presented in the session of 20 November 1915.

The far reaching ideas and the formation of novel concepts by means of which Mie constructs his electrodynamics, and the prodigious problems raised by Einstein, as well as his ingeniously conceived methods of solution, have opened new paths for the investigation into the foundations of physics.

In the following — in the sense of the axiomatic method — I would like to develop /*essentially* from three simple axioms a ~~new~~ system of basic equations of physics, of ideal beauty, containing, I believe, the solution of the problems presented. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.

Let w_s $(s = 1, 2, 3, 4)$ be any coordinates labeling the world's points essentially uniquely — the so-called world parameters. The quantities characterizing the events at w_s shall be:

1) The ten gravitational potentials /*first introduced by Einstein* $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ having the character of a symmetric tensor with respect to arbitrary transformation of the world parameter w_s;

2) The four electrodynamic potentials q_s having the character of a vector in the same sense.

Physical processes do not proceed in an arbitrary way, rather they are governed by the following two axioms: |

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

[2] Axiom I (Mie's[1] axiom of the world function): *The law governing physical processes is determined through a world function* H, *that contains the following arguments*:

$$g_{\mu\nu}, \qquad g_{\mu\nu l} = \frac{\partial g_{\mu\nu}}{\partial w_l}, \qquad g_{\mu\nu lk} = \frac{\partial^2 g_{\mu\nu}}{\partial w_l \partial w_k}, \tag{1}$$

$$q_s, \qquad q_{sl} = \frac{\partial q_s}{\partial w_l} \qquad (l, k = 1, 2, 3, 4) \tag{2}$$

where the variation of the integral

$$\int H \sqrt{g}\, d\tau;$$

$$(g = |g_{\mu\nu}|, \quad d\tau = dw_1 dw_2 dw_3 dw_4)$$

must vanish for each of the 14 potentials $g_{\mu\nu}, q_s$.

Clearly the arguments (1) can be replaced by the arguments

$$g^{\mu\nu}, \qquad g^{\mu\nu}_l = \frac{\partial g^{\mu\nu}}{\partial w_l}, \qquad g^{\mu\nu}_{lk} = \frac{\partial^2 g^{\mu\nu}}{\partial w_l \partial w_k} \tag{3}$$

where $g^{\mu\nu}$ is the subdeterminant of the determinant g with respect to its element $g_{\mu\nu}$, divided by g.

Axiom II[2] (axiom of general invariance): *The world function* H *is invariant with respect to an arbitrary transformation of the world parameters* w_s.

Axiom II is the simplest mathematical expression of the demand that the interlinking of the potentials $g_{\mu\nu}, q_s$ is by itself entirely independent of the way one chooses to identify the world's points by means of world parameters.

The guiding motive for setting up ~~my~~ *the* theory is given by the following theorem, the proof of which I shall present elsewhere.

Theorem I. If J is an invariant under arbitrary transformations of the four world
[3] parameters, containing n quantities and their derivatives, | and if one forms from

$$\delta \int J \sqrt{g}\, d\tau = 0$$

---

1 Mie's world functions do not contain exactly these arguments; in particular the usage of the arguments (2) goes back to Born. However, what is characteristic of Mie's electrodynamics is the introduction and use of such a world function in Hamilton's principle.

2 Orthogonal invariance was already postulated by Mie. In the axiom II established above, Einstein's basic idea fundamental[2] of general covariance finds its simplest expression, even if Hamilton's principle plays only a subsidiary role with Einstein, and his functions H are by no means invariants, and also do not contain the electric potentials.

the n variational equations of Lagrange with respect to each of the n quantities, then in this invariant system of n differential equations for the n quantities there are always four that are a consequence of the remaining $n-4$ — in the sense that, among the n differential equations and their total derivatives there are always four linear and mutually independent combinations that are satisfied identically.

Concerning the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$, $g_{kl}^{\mu\nu}$ as in (4) and subsequent formulas, let us note once for all that, due to the symmetry in μ, ν on the one hand and in k, l on the other, the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$, according as $\mu = \nu$ resp. $\mu \neq \nu$, further the differential quotients with respect to $g_{kl}^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$ resp. $\frac{1}{4}$, according as $\mu = \nu$ and $k = l$ resp. $\mu = \nu$ and $k \neq l$ or $\mu \neq \nu$ and $k = l$ resp. $\mu \neq \nu$ and $k \neq l$.

Axiom I implies first for the ten gravitational potentials $g^{\mu\nu}$ the ten Lagrangian differential equations

$$\frac{\partial \sqrt{g} H}{\partial g^{\mu\nu}} = \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g} H}{\partial g_k^{\mu\nu}} - \sum_{k,l} \frac{\partial^2}{\partial w_k \partial w_l} \frac{\partial \sqrt{g} H}{\partial g_{kl}^{\mu\nu}}, \qquad (\mu, \nu = 1, 2, 3, 4) \tag{4}$$

and secondly for the four electrodynamic potentials q_s the four Lagrangian differential equations

$$\frac{\partial \sqrt{g} H}{\partial q_h} = \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g} H}{\partial q_{hk}}, \qquad (h = 1, 2, 3, 4). \tag{5}$$

Let us call equations (4) the fundamental equations of gravitation, and equations (5) the fundamental electrodynamic equations, or generalized Maxwell equations. Due to the theorem stated above, the four equations (5) can be viewed as a consequence of equations (4), that is, because of that mathematical theorem we can immediately assert the claim *that in the sense explained above electrodynamic phenomena are effects of gravitation*. I regard this insight as the simple and very surprising solution of the problem of Riemann, who was the first to search for a theoretical connection between gravitation and light.

Since our mathematical theorem shows us that the axioms I and II considered so far can produce only ten essentially independent equations; and since, on the other hand, | if general invariance is maintained, more than ten essentially independent [4] equations for the 14 potentials $g_{\mu\nu}, q_s$ are not at all possible; therefore—provided that we want to retain the determinate character of the basic equation of physics corresponding to Cauchy's theory of differential equations— the demand for four further non-invariant equations in addition to (4) and (5) is imperative. In order to arrive at these equations, I first put up a definition of the concept of energy.

To this end we polarize $g^{\mu\nu}$ in the invariant H by an arbitrary contragredient tensor $h^{\mu\nu}$ and thus form the expression

$$J^{(h)} = \sum_{\mu,\nu} \frac{\partial H}{\partial g^{\mu\nu}} h^{\mu\nu} + \sum_{\mu,\nu,k} \frac{\partial H}{\partial g_k^{\mu\nu}} h_k^{\mu\nu} + \sum_{\mu,\nu,k,l} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} h_{kl}^{\mu\nu},$$

where the abbreviation

$$h_k^{\mu\nu} = \frac{\partial h^{\mu\nu}}{\partial w_k}, \qquad h_{kl}^{\mu\nu} = \frac{\partial^2 h^{\mu\nu}}{\partial w_k \partial w_l}$$

has been used. Since polarization is an invariant process, $J^{(h)}$ is an invariant. Now we treat the expression $\sqrt{g}J^{(h)}$ in the same way as an integrand of a variational problem in the calculus of variations, when one wants to integrate by parts; thus we obtain the following identity:

$$\sqrt{g}J^{(h)} = -\sum_{\mu,\nu} H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} h^{\mu\nu} + \sum_{\mu,\nu} [\sqrt{g}H]_{\mu\nu} h^{\mu\nu} + D^{(h)}, \tag{6}$$

where we have put

$$[\sqrt{g}H]_{\mu\nu} = \frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l} \frac{\partial^2}{\partial w_k \partial w_l} \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}}$$

and

$$D^{(h)} = \sum_{\mu,\nu,k} \frac{\partial}{\partial w_k} \left(\frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} h^{\mu\nu} \right) + \sum_{\mu,\nu,k,l} \frac{\partial}{\partial w_k} \left(\frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} h_l^{\mu\nu} \right) - \sum_{\mu,\nu,k,l} \frac{\partial}{\partial w_l} \left(h^{\mu\nu} \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} \right) \tag{7}$$

as abbreviations. The expression $[\sqrt{g}H]_{\mu\nu}$ is nothing but the Lagrangian variational derivative of $\sqrt{g}H$ with respect to $g_{\mu\nu}$, which yields the gravitational equations (4) when it is put equal to zero,

$$[\sqrt{g}H]_{\mu\nu} = 0 \tag{8}$$

[5] | and the expression $D^{(h)}$ is a sum of differential quotients, so it has the character of a pure divergence.

Now we use the easily proved fact that, if p^j $(j = 1, 2, 3, 4)$ is an arbitrary contravariant vector, the expression

$$p^{\mu\nu} = \sum_s (g_s^{\mu\nu} p^s - g^{\mu s} p_s^{\nu} - g^{\nu s} p_s^{\mu}), \qquad \left(p_s^j = \frac{\partial p^j}{\partial w_s} \right)$$

represents a symmetric contravariant tensor.

If we substitute in the invariant expression $J^{(h)}$ instead of $h^{\mu\nu}$ the special contravariant tensor $p^{\mu\nu}$, there arises again an invariant expression, namely

$$J^{(p)} = \sum_{\mu,\nu} \frac{\partial H}{\partial g^{\mu\nu}} p^{\mu\nu} + \sum_{\mu,\nu,k} \frac{\partial H}{\partial g^{\mu\nu}_k} p^{\mu\nu}_k + \sum_{\mu,\nu,k,l} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} p^{\mu\nu}_{kl},$$

where the abbreviations

$$p^{\mu\nu}_k = \frac{\partial p^{\mu\nu}}{\partial w_k}, \qquad p^{\mu\nu}_{kl} = \frac{\partial^2 p^{\mu\nu}}{\partial w_k \partial w_l}$$

have been used. Now we treat the expression $\sqrt{g} J^{(p)}$ in the same way as an integrand of a variational problem in the calculus of variations, when one wants to integrate by parts — but in such a way that in this procedure the first differential quotient p^j_s of the p^j always remain unchanged, and only the second and third derivatives of the p^j are included in the divergence; and moreover so that the auxiliary expressions become invariant with respect to linear transformation

$$\begin{aligned} E = {} & \sum \left(H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} g^{\mu\nu}_s + \sqrt{g} \frac{\partial H}{\partial g^{\mu\nu}} g^{\mu\nu}_s + \sqrt{g} \frac{\partial H}{\partial g^{\mu\nu}_k} g^{\mu\nu}_{sk} + \sqrt{g} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} g^{\mu\nu}_{skl} \right) p \Big/^s_s \\ & - \sum (g^{\mu s} p^\nu_s + g^{\nu s} p^\mu_s) [\sqrt{g} H]_{\mu\nu} \\ & + \sum \left(\frac{\partial \sqrt{g} H}{\partial g^{\mu\nu}_k} - g^{\mu\nu}_s + \frac{\partial \sqrt{g} H}{\partial g^{\mu\nu}_{kl}} - g^{\mu\nu}_{sl} - g^{\mu\nu}_s \frac{\partial}{\partial w_l} \frac{\partial \sqrt{g} H}{\partial g^{\mu\nu}_{kl}} \right) p^s_k; \end{aligned} \tag{9}$$

we thus obtain the following identity:

$$\sqrt{g} J^{(p)} = -\sum_{\mu,\nu} H \frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} p^{\mu\nu} + E + D^{(p)}, \tag{10}$$

where we have put | [6]

$$\begin{aligned} D^{(p)} = {} & \sum \Bigg\{ - \frac{\partial}{\partial w_k} \left(\sqrt{g} \frac{\partial H}{\partial g^{\mu\nu}_k} (g^{\mu s} p^\nu_s + g^{\nu s} p^\mu_s) \right) \\ & + \frac{\partial}{\partial w_k} \left((p^\nu_s g^{\mu s} + p^\mu_s g^{\nu s}) \frac{\partial}{\partial w_l} \left(\sqrt{g} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} \right) \right) \\ & + \frac{\partial}{\partial w_l} \left(\sqrt{g} \frac{\partial H}{\partial g^{\mu\nu}_{kl}} \left(\frac{\partial p^{\mu\nu}}{\partial w_k} - g^{\mu\nu}_{sk} p^s \right) \right) \Bigg\} \end{aligned}$$

as an abbreviation.[3] The expression E is invariant under linear transformation and with respect to the vector p^j it has the form

$$E = \sum_s e_s p^s + \sum_{s,l} e_s^l p_l^s,$$

where from (10) e_s and e_s^l are well-defined expressions. In particular it turns out, as one can see, that:

$$e_s = \frac{d^{(g)}\sqrt{g}H}{dw_s}; \tag{11}$$

where the differentiation denoted by $d^{(g)}$ is total with respect to w_s, but to be performed in such a way that the electromagnetic potentials q_s remain unaffected.

Call the expression E the energy form. To justify this designation, I prove two properties that the energy form enjoys.

If we substitute the tensor $p^{\mu\nu}$ for $h^{\mu\nu}$ in identity (6) then, together with (9) it follows, provided the gravitational equations (8) are satisfied:

$$E = (D^{(h)})_{h=p} - D^{(p)} \tag{12}$$

or

$$E = \sum \left\{ \frac{\partial}{\partial w_k}\left(\sqrt{g}\frac{\partial H}{\partial g_k^{\mu\nu}} g_s^{\mu\nu} p^s\right) - \frac{\partial}{\partial w_k}\left(\frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}}\right) g_s^{\mu\nu} p^s\right) + \frac{\partial}{\partial w_l}\left(\sqrt{g}\frac{\partial H}{\partial g_{kl}^{\mu\nu}} g_{sk}^{\mu\nu} p^s\right) \right\}, \tag{13}$$

that is, we have the proposition:

Proposition 1: In virtue of the gravitational equations the energy form E becomes a sum of differential quotients with respect to w_s, that is, it acquires the character of a divergence.

Had we gone a step further in the above treatment of the expression $\sqrt{g}J^{(p)}$, that led to (9) and converted in the usual way of the variational calculus also the first differential quotient p_s^j of the p^j then the expression containing the p^j alone would |

[eq (14) missing][4] (14)

[7] This theorem shows that the divergence equation corresponding to the energy theorem of the old theory

$$\sum_l \frac{\partial e_s^l}{\partial w_l} = 0 \tag{15}$$

holds if and only if the four quantities e_s vanish, that is if the following equations hold

$$\frac{d^{(g)}\sqrt{g}H}{dw_s} = 0. \tag{16}$$

After these preliminaries I now put down the following axiom:

Axiom III (axiom of space and time). *The spacetime coordinates are those special world parameters for which the energy theorem* (15) *is valid.*

According to this axiom, space and time in reality provide a special labeling of the world's points such that the energy theorem holds.

Axiom III implies the existence of equations (16): these four differential equations (16) complete the gravitational equations (4) to give a system of 14 equations for the 14 potentials $g_{\mu\nu}, q_s$, *the system of fundamental equations of physics*. Because of the agreement in number between equations and potentials to be determined, the principle of causality for physical processes is also guaranteed, revealing to us the closest connection between the energy theorem and the principle of causality, since each presupposes the other. To the transition from one spacetime reference system to another one corresponds the transformation of the energy form from one so-called "normal form"

$$E = \sum_{s,l} e_s^l p_l^s$$

to another normal form. |

$$[\text{eq. (17) missing: } H = K + L.]^{[4]} \tag{17}$$

Because K depends only on $g^{\mu\nu}$, $g_s^{\mu\nu}$, $g_{kl}^{\mu\nu}$, therefore in ansatz (17), due to (13), [8] the energy E can be expressed solely as a function of the gravitational potentials $g^{\mu\nu}$ and their derivatives, provided L is assumed to depend not on $g_s^{\mu\nu}$, but only on $g^{\mu\nu}$, q_s, q_{sk}. On this assumption, which we shall always make in the following, the definition of the energy (10) yields the expression

$$E = E^{(g)} + E^{(e)}, \tag{18}$$

where the "gravitational energy" $E^{(g)}$ depends only on $g^{\mu\nu}$ and their derivatives, and the "electrodynamic energy" $E^{(e)}$ takes the form

$$E^{(e)} = \sum_{\mu,\nu,s} \frac{\partial\sqrt{g}L}{\partial g^{\mu\nu}}(g_s^{\mu\nu}p^s - g^{\mu s}p_s^{\nu} - g^{\nu s}p_s^{\mu}), \tag{19}$$

which proves to be *a general invariant multiplied by* $\sqrt{g}$.

To proceed we use two mathematical theorems, which say the following:

Theorem II. If J is an invariant depending on $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{kl}^{\mu\nu}$, q_s, q_{sk}, then the following is always identically true in all arguments and for every arbitrary contravariant vector p^s:

$$\sum_{\mu,\nu,l,k}\left(\frac{\partial J}{\partial g^{\mu\nu}}\Delta g^{\mu\nu}+\frac{\partial J}{\partial g^{\mu\nu}_{l}}\Delta g^{\mu\nu}_{l}+\frac{\partial J}{\partial g^{\mu\nu}_{kl}}\Delta g^{\mu\nu}_{kl}\right)$$

$$+\sum_{s,k}\left(\frac{\partial J}{\partial q_s}\Delta q_s+\frac{\partial J}{\partial q_{sk}}\Delta q_{sk}\right)=0;$$

where |

$$\Delta g^{\mu\nu}=\sum_{m}(g^{\mu m}p^{\nu}_{m}+g^{\nu m}p^{\mu}_{m}),$$

$$\Delta g^{\mu\nu}_{l}=-\sum_{m}g^{\mu\nu}_{m}p^{m}_{l}+\frac{\partial\Delta g^{\mu\nu}}{\partial w_l},$$

[9]
$$\Delta g^{\mu\nu}_{lk}=-\sum_{m}(g^{\mu\nu}_{m}p^{m}_{lk}+g^{\mu\nu}_{lm}p^{m}_{k}+g^{\mu\nu}_{km}p^{m}_{l})+\frac{\partial^2\Delta g^{\mu\nu}}{\partial w_l\partial w_k},$$

$$\Delta q_s=-\sum_{m}q_m p^{m}_{s},$$

$$\Delta q_{sk}=-\sum_{m}q_{sm}p^{m}_{k}+\frac{\partial\Delta q_s}{\partial w_k}.$$

Theorem III. If J is an invariant depending only on the $g^{\mu\nu}$ and their derivatives and if, as above, the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$ are denoted by $[\sqrt{g}J]_{\mu\nu}$ then the expression — in which $h^{\mu\nu}$ is understood to be any contravariant tensor —

$$\frac{1}{\sqrt{g}}\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}h^{\mu\nu}$$

represents an invariant; if in this sum we substitute in place of $h^{\mu\nu}$ the particular tensor $p^{\mu\nu}$ and write

$$\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}p^{\mu\nu}=\sum_{s,l}(i_s p^{s}+i^{l}_{s}p^{s}_{l}),$$

where then the expressions

$$i_s=\sum_{\mu,\nu}[\sqrt{g}J]_{\mu\nu}g^{\mu\nu}_{s},$$

$$i^{l}_{s}=-2\sum_{\mu}[\sqrt{g}J]_{\mu s}g^{\mu l}$$

depend only on the $g^{\mu\nu}$ and their derivatives, then we have

$$i_s=\sum_{l}\frac{\partial i^{l}_{s}}{\partial w_l} \tag{20}$$

in the sense, that this equation is identically fulfilled for all arguments, that is for the $g^{\mu\nu}$ and their derivatives.

Now we apply Theorem II to the invariant L and obtain

$$\sum_{\mu,\nu,m} \frac{\partial L}{\partial g^{\mu\nu}}(g^{\mu m}p_m^{\nu} + g^{\nu m}p_m^{\mu}) - \sum_{s,m} \frac{\partial L}{\partial q_s} q_m p_s^m \tag{21}$$

$$- \sum_{s,k,m} \frac{\partial L}{\partial q_{sk}}(q_{sm}p_k^m + q_{mk}p_s^m + q_m p_{sk}^m) = 0.$$

Equating to zero the coefficient of p_{sk}^m produces the equation

$$\left(\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}}\right) q_m = 0$$

| or [10]

$$\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0, \tag{22}$$

that is, the derivatives of the electrodynamic potentials q_s occur only in the combinations

$$M_{ks} = q_{sk} - q_{ks}.$$

Thus we learn that under our assumptions the invariant L depends, other than on the potentials $g_{\mu\nu}, q_s$, only on the components of the skew symmetric invariant tensor

$$M = (M_{ks}) = \mathrm{rot}(q_s),$$

that is, of the so-called electromagnetic six vector. *This result here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*

If we put the coefficient of p_m^{ν} on the left of identity (21) equal to zero, we obtain, using (22)

$$2\sum_{\mu} \frac{\partial L}{\partial g^{\mu\nu}} g^{\mu m} - \frac{\partial L}{\partial q_m} q_{\nu} - \sum_{s} \frac{\partial L}{\partial M_{ms}} M_{\nu s} = 0, \qquad (\mu = 1, 2, 3, 4). \tag{23}$$

This equation admits an important transformation of the electromagnetic energy. Namely, the part of $E^{(e)}$ multiplied by p_m^{ν} in (19) becomes due to (23):

$$-2\sum_{\mu} \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} g^{\mu m} = \sqrt{g}\left\{ L\delta_{\nu}^m - \frac{\partial L}{\partial q_m} q_{\nu} - \sum_{s} \frac{\partial L}{\partial M_{ms}} M_{\nu s} \right\}, \tag{24}$$

$$(\mu = 1, 2, 3, 4) \quad (\delta_{\nu}^{\mu} = 0, \ \mu \neq \nu, \quad \delta_{\mu}^{\mu} = 1).$$

If one subjects the expression on the right to the limit

$$\begin{aligned} g_{\mu\nu} &= 0, \qquad (\mu \neq \nu) \\ g_{\mu\mu} &= 1, \end{aligned} \tag{25}$$

then this limit agrees exactly with what Mie has established in his electrodynamics: Mie's electromagnetic energy tensor is nothing but the generally invariant tensor that results from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in the limit (25) — a circumstance that gave me the first hint of the necessary close connection between Einstein's general relativity theory and Mie's electrodynamics, and which convinced me of the correctness of the theory here developed. |

[11] It only remains to show directly from assumption (17) how the generalized Maxwell equations (5) developed above are a consequence of the gravitational equations (4) in the sense given above.

By use of the notation just introduced for the variational derivatives with respect to the $g^{\mu\nu}$ the gravitational equations acquire the form, due to (17)

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial \sqrt{g}L}{\partial g^{\mu\nu}} = 0. \tag{26}$$

If we further denote in general the variational derivatives of $\sqrt{g}J$ with respect to the electrodynamic potential q_h by

$$[\sqrt{g}J]_h = \frac{\partial \sqrt{g}J}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}J}{\partial q_{hk}},$$

then the electrodynamic equations take the form, due to (17)

$$[\sqrt{g}L]_h = 0. \tag{27}$$

Since K is an invariant that depends only on $g^{\mu\nu}$ and its derivatives, then by theorem III equation (20) holds identically with

$$i_s = \sum_{\mu,\nu} [\sqrt{g}K]_{\mu\nu} g^{\mu\nu}_s \tag{28}$$

and

$$i^l_s = -2 \sum_\mu [\sqrt{g}K]_{\mu s} g^{\mu l}, \qquad (\mu = 1, 2, 3, 4). \tag{29}$$

Because of (26) and (29) the left side of (24) equals $-i^m_\nu$. By differentiation with respect to w_m and summation over m we obtain because of (20)

$$i_\nu = \sum_m \frac{\partial}{\partial w_m} \left(-\sqrt{g}L\delta^m_\nu + \frac{\partial \sqrt{g}L}{\partial q_m} q_\nu + \sum_s \frac{\partial \sqrt{g}L}{\partial M_{sm}} M_{s\nu} \right)$$

$$= -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \left\{ q_\nu \frac{\partial}{\partial w_m}\left([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\right) + q_{\nu m}\left([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\right)\right\} + \sum_s \left([\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}\right) M_{s\nu} + \sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m},$$

since of course

$$\frac{\partial\sqrt{g}L}{\partial q_m} = [\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}$$

l and [12]

$$-\sum_m \frac{\partial}{\partial w_m}\frac{\partial\sqrt{g}L}{\partial q_{sm}} = [\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}.$$

Now we take into account that because of (22) we have

$$\sum_{m,s} \frac{\partial^2}{\partial w_m \partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}} = 0,$$

and thus obtain after suitably collecting terms

$$i_\nu = -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \left(q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m + M_{m\nu}[\sqrt{g}L]_m\right) + \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} + \sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}. \tag{30}$$

On the other hand we have

$$-\frac{\partial\sqrt{g}L}{\partial w_\nu} = -\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial g^{sm}} g_\nu^{sm} - \sum_m \frac{\partial\sqrt{g}L}{\partial q_m} q_{m\nu} - \sum_{m,s} \frac{\partial\sqrt{g}L}{\partial q_{ms}}\frac{\partial q_{ms}}{\partial w_\nu}.$$

Due to (26) and (28) the first terms on the right is nothing else but i_ν. The last term on the right proves to be equal and opposite to the last term on the right of (30); for we have

$$\sum_{s,m} \frac{\partial\sqrt{g}L}{\partial M_{sm}}\left(\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu}\right) = 0, \tag{31}$$

since the expression

$$\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu} = \frac{\partial^2 q_\nu}{\partial w_s \partial w_m} - \frac{\partial^2 q_s}{\partial w_\nu \partial w_m} - \frac{\partial^2 q_m}{\partial w_\nu \partial w_s}$$

turns out to be symmetric in s, m, and the first factor under the summation sign in (31) skew symmetric in s, m.

Therefore (30) implies the equations

$$\sum_m \left(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m \right) = 0; \tag{32}$$

that is, from the gravitational equations (4) there follow indeed the four linearly independent combinations (32) of the basic electrodynamic equations (5) and their first derivatives. *This is the entire mathematical expression of the general claim made above about the character of electrodynamics as an epiphenomenon of gravitation.* |

[13] According to our assumption L should not depend on the derivatives of the $g^{\mu\nu}$, therefore L must be a function of certain four general invariants, which correspond to the special orthogonal invariants reported by Mie, and of which the two simplest ones are these:

$$Q = \sum_{k,l,m,n} M_{mn} M_{lk} g^{mk} g^{nl}$$

and

$$q = \sum_{k,l} q_k q_l g^{kl}.$$

The simplest and most straightforward ansatz for L, considering the structure of K, is also that which corresponds to Mie's electrodynamics, namely

$$L = \alpha Q + f(q)$$

or, following Mie even more closely:

$$L = \alpha Q + \beta q^3,$$

where $f(q)$ denotes any function of q, and α, β are constants.

As one can see, the few simple assumptions expressed in axioms I, II, III suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense called for by Einstein, but I am also convinced that through the basic equations established here the most intimate, hitherto hidden processes in the interior of atoms will receive an explanation; and in particular that generally a reduction of all physical constants to mathematical constants must be possible—whereby the possibility approaches that physics in principle becomes a science of the type of geometry: surely the highest glory of the axiomatic method, which, as we have seen, here takes into its service the

powerful instruments of analysis, namely the calculus of variations and the theory of invariants.

EDITORIAL NOTES

[1] The following is a translation of the proofs of Hilbert's first paper on the foundations of physics, which are preserved at Göttingen in SUB Cod. Ms. 634. These proofs comprise 13 pages and are complete, apart from the fact that roughly the upper quarter of two pages (7 and 8) is cut off. The proofs are dated "submitted on 20 November 1915." The Göttingen copy bears a printer's stamp dated 6 December 1915 and is marked in Hilbert's own hand "First proofs of my first note." In addition, the proofs carry several marginal notes in Hilbert's hand, which are shown here in superscript italics. In contrast to the other source papers in these volumes, this proof version of Hilbert's paper has been formatted as far as possible to recreate the original so that the author's hand-written notes are evident. This paper was later published in a converted version in *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse*. 1915. Issue 8, p 395–407, (1. correction).

[2] The word "fundamental" should appear before "basic". It is written correctly in the printed version.

[3] The superscript $^{\nu}$ in the first occurrence of p_s^{ν} in this equation is missing in the original.

[4] For more detailed information on the missing piece of this document, see "Einstein Equations and Hilbert Actions …" (in this volume).

DAVID HILBERT

THE FOUNDATIONS OF PHYSICS (FIRST COMMUNICATION)

Originally published as "Die Grundlagen der Physik. (Erste Mitteilung)" in Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse. 1916. Issue 8, p. 395–407. Presented in the session of 20 November 1915.

The vast problems posed by Einstein[1] as well as his ingeniously conceived methods of solution, and the far-reaching ideas and formation of novel concepts by means of which Mie[2] constructs his electrodynamics, have opened new paths for the investigation into the foundations of physics.

In the following—in the sense of the axiomatic method—I would like to develop, essentially from two simple axioms, a new system of basic equations of physics, of ideal beauty and containing, I believe, *simultaneously* the solution to the problems of Einstein and of Mie. I reserve for later communications the detailed development and particularly the special application of my basic equations to the fundamental questions of the theory of electricity.

Let w_s $(s = 1, 2, 3, 4)$ be any coordinates labeling the world's points essentially uniquely—the so-called world parameters (most general spacetime coordinates). The quantities characterizing the events at w_s shall be:

1. The ten gravitational potentials $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ first introduced by Einstein, having the character of a symmetric tensor with respect to an arbitrary transformation of the world parameters w_s;
2. The four electrodynamic potentials q_s having the character of a vector in the same sense. |

Physical processes do not proceed in an arbitrary way, rather they are governed by [396]
the following two axioms:

1 *Sitzungsber. d. Berliner Akad.* 1914, 1030; 1915, 778, 799, 831, 844.
2 *Ann. d. Phys.* 1912, Vol. 37, 511; Vol. 39, 1; 1913, vol. 40, 1.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

Axiom I (Mie's axiom of the world function[3]): *The law governing physical processes is determined through a world function* H, *that contains the following arguments:*

$$g_{\mu\nu}, \qquad g_{\mu\nu l} = \frac{\partial g_{\mu\nu}}{\partial w_l}, \qquad g_{\mu\nu lk} = \frac{\partial^2 g_{\mu\nu}}{\partial w_l \partial w_k}, \tag{1}$$

$$q_s, \qquad q_{sl} = \frac{\partial q_s}{\partial w_l} \qquad (l, k = 1, 2, 3, 4), \tag{2}$$

where the variation of the integral

$$\int H \sqrt{g} d\omega$$

$$(g = |g_{\mu\nu}|, \quad d\omega = dw_1 dw_2 dw_3 dw_4)$$

must vanish for each of the fourteen potentials $g_{\mu\nu}, q_s$.

Clearly the arguments (1) can be replaced by the arguments [1]

$$g^{\mu\nu}, \qquad g^{\mu\nu}_l = \frac{\partial g^{\mu\nu}}{\partial w_l}, \qquad g^{\mu\nu}_{lk} = \frac{\partial^2 g^{\mu\nu}}{\partial w_l \partial w_k}, \tag{3}$$

where $g^{\mu\nu}$ is the subdeterminant of the determinant g with respect to its element $g_{\mu\nu}$, divided by g.

Axiom II (axiom of general invariance[4]): *The world function* H *is invariant with respect to an arbitrary transformation of the world parameters* w_s.

Axiom II is the simplest mathematical expression of the demand that the interlinking of the potentials $g_{\mu\nu}, q_s$ is by itself entirely independent of the way one chooses to label the world's points by means of world parameters.

The guiding motive for constructing my theory is provided by the following theorem, the proof of which I shall present elsewhere. |

[397] Theorem I. If J is an invariant under arbitrary transformation of the four world parameters, containing n quantities and their derivatives, and if one forms from

$$\delta \int J g d\omega = 0$$

the n variational equations of Lagrange with respect to those n quantities, then in this invariant system of n differential equations for the n quantities there are always

---

3 Mie's world functions do not contain exactly these arguments; in particular the usage of the arguments (2) goes back to Born. However, what is characteristic of Mie's electrodynamics is precisely the introduction and use of such a world function in Hamilton's principle.

4 Orthogonal invariance was already postulated by Mie. In the axiom II formulated above, Einstein's fundamental basic idea of general invariance finds its simplest expression, even if Hamilton's principle plays only a subsidiary role with Einstein, and his functions H are by no means general invariants, and also do not contain the electric potentials.

four that are a consequence of the remaining $n-4$ —in this sense, that among the n differential equations and their total derivatives there are always four linear and mutually independent combinations that are satisfied identically.

Concerning the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$, $g_{kl}^{\mu\nu}$ occurring in (4) and subsequent formulas, let us note once for all that, due to the symmetry in μ, ν on the one hand and in k, l on the other, the differential quotients with respect to $g^{\mu\nu}$, $g_k^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$, according as $\mu = \nu$ resp. $\mu \neq \nu$, further the differential quotients with respect to $g_{kl}^{\mu\nu}$ are to be multiplied by 1 resp. $\frac{1}{2}$ resp. $\frac{1}{4}$, according as $\mu = \nu$ and $k = l$ resp. $\mu = \nu$ and $k \neq l$ or $\mu \neq \nu$ and $k = l$ resp. $\mu \neq \nu$ and $k \neq l$.

Axiom I implies first for the ten gravitational potentials $g^{\mu\nu}$ the ten Lagrangian differential equations

$$\frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} - \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} + \sum_{k,l} \frac{\partial^2}{\partial w_k \partial w_l} \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} = 0, \qquad (\mu, \nu = 1, 2, 3, 4) \tag{4}$$

and secondly for the four electrodynamic potentials q_s the four Lagrangian differential equations

$$\frac{\partial \sqrt{g}H}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k} \frac{\partial \sqrt{g}H}{\partial q_{hk}} = 0, \qquad (h = 1, 2, 3, 4). \tag{5}$$

We denote the left sides of the equations (4), (5) respectively by

$$[\sqrt{g}H]_{\mu\nu}, \qquad [\sqrt{g}H]_h$$

for short.

Let us call equations (4) the fundamental equations of gravitation, and equations (5) the fundamental electrodynamic equations, or generalized Maxwell equations. Due to the theorem stated above, the four equations (5) can be viewed as a consequence of equations (4), that is, because of that mathematical theorem we can directly make the claim *that in the sense as explained the electrodynamic phenomena are effects of gravitation*. I regard this | insight as the simple and very surprising solu- [398]
tion of the problem of Riemann, who was the first to search for a theoretical connection between gravitation and light.

In the following we use the easily proved fact that, if p^j $(j = 1, 2, 3, 4)$ is an *arbitrary* contravariant vector, the expression

$$p^{\mu\nu} = \sum_s (g_s^{\mu\nu} p^s - g^{\mu s} p_s^{\nu} - g^{\nu s} p_s^{\mu}), \qquad \left(p_s^j = \frac{\partial p^j}{\partial w_s}\right)$$

represents a symmetric contravariant tensor, and the expression

$$p_l = \sum_s (q_{ls} p^s + q_s p_l^s)$$

represents a covariant vector.

To proceed we establish two mathematical theorems, which express the following:

Theorem II. If J is an invariant depending on $g^{\mu\nu}$, $g_l^{\mu\nu}$, $g_{kl}^{\mu\nu}$, q_s, q_{sk}, then the following is always identically true in all arguments and for every arbitrary contravariant vector p^s: [2]

$$\sum_{\mu,\nu,l,k}\left(\frac{\partial J}{\partial g^{\mu\nu}}\Delta g^{\mu\nu}+\frac{\partial J}{\partial g_l^{\mu\nu}}\Delta g_l^{\mu\nu}+\frac{\partial J}{\partial g_{kl}^{\mu\nu}}\Delta g_{kl}^{\mu\nu}\right)$$
$$+\sum_{s,k}\left(\frac{\partial J}{\partial q_s}\Delta q_s+\frac{\partial J}{\partial q_{sk}}\Delta q_{sk}\right)=0,$$

where

$$\Delta g^{\mu\nu}=\sum_m(g^{\mu m}p_m^{\nu}+g^{\nu m}p_m^{\mu}),$$
$$\Delta g_l^{\mu\nu}=-\sum_m g_m^{\mu\nu}p_l^m+\frac{\partial\Delta g^{\mu\nu}}{\partial w_l},$$
$$\Delta g_{lk}^{\mu\nu}=-\sum_m(g_m^{\mu\nu}p_{lk}^m+g_{lm}^{\mu\nu}p_k^m+g_{km}^{\mu\nu}p_l^m)+\frac{\partial^2\Delta g^{\mu\nu}}{\partial w_l\partial w_k},$$
$$\Delta q_s=-\sum_m q_m p_s^m,$$
$$\Delta q_{sk}=-\sum_m q_{sm}p_k^m+\frac{\partial\Delta q_s}{\partial w_k}.$$

This theorem II can also be formulated as follows:

If J is an invariant and p^s and arbitrary vector as above, then the identity holds

$$\sum_s\frac{\partial J}{\partial w_s}p^s=PJ, \tag{6}$$

[399] | where we have put

$$P=P_g+P_q,$$

with

$$P_g=\sum_{\mu,\nu,l,k}\left(p^{\mu\nu}\frac{\partial}{\partial g^{\mu\nu}}+p_l^{\mu\nu}\frac{\partial}{\partial g_l^{\mu\nu}}+p_{lk}^{\mu\nu}\frac{\partial}{\partial g_{lk}^{\mu\nu}}\right)$$
$$P_q=\sum_{l,k}\left(p_l\frac{\partial}{\partial q_l}+p_{lk}\frac{\partial}{\partial q_{lk}}\right),$$

and used the abbreviations:

$$p_k^{\mu\nu} = \frac{\partial p^{\mu\nu}}{\partial w_k}, \qquad p_{kl}^{\mu\nu} = \frac{\partial^2 p^{\mu\nu}}{\partial w_k \partial w_l}, \qquad p_{lk} = \frac{\partial p_l}{\partial w_k}.$$

The proof of (6) follows easily; for this identity is obviously correct if p^s is a constant vector, and from this it follows in general because of its invariance.

Theorem III. If J is an invariant depending *only* on $g^{\mu\nu}$ and their derivatives, and if, as above, the variational derivatives of $\sqrt{g}J$ with respect to $g^{\mu\nu}$ are denoted by $[\sqrt{g}J]_{\mu\nu}$ then the expression—where $h^{\mu\nu}$ is understood to be any contravariant tensor—

$$\frac{1}{\sqrt{g}} \sum_{\mu,\nu} [\sqrt{g}J]_{\mu\nu} h^{\mu\nu}$$

represents an invariant; if we substitute in this sum in place of $h^{\mu\nu}$ the particular tensor $p^{\mu\nu}$ and write

$$\sum_{\mu,\nu} [\sqrt{g}J]_{\mu\nu} p^{\mu\nu} = \sum_{s,l} (i_s p^s + i_s^l p_l^s),$$

where then the expressions

$$i_s = \sum_{\mu,\nu} [\sqrt{g}J]_{\mu\nu} g_s^{\mu\nu},$$
$$i_s^l = -2 \sum_{\mu} [\sqrt{g}J]_{\mu s} g^{\mu l}$$

depend only on the $g^{\mu\nu}$ and their derivatives, then we have

$$i_s = \sum_l \frac{\partial i_s^l}{\partial w_l} \tag{7}$$

in the sense that this equation is satisfied identically for all arguments, that is for the $g^{\mu\nu}$ and their derivatives.

For the proof we consider the integral

$$\int J\sqrt{g}\, d\omega, \qquad d\omega = dw_1 dw_2 dw_3 dw_4$$

to be taken over a finite piece of the four dimensional world. | Further, let p^s be a [400]
vector that vanishes together with its derivatives on the three dimensional surface of that piece of the world. Due to $P = P_g$ the last formula of the next page implies

$$P_g(\sqrt{g}J) = \sum_s \frac{\partial \sqrt{g} J p^s}{\partial w_s};$$

this results in

$$\int P_g(J\sqrt{g})d\omega = 0$$

and due to the way the Lagrangian derivative is formed we accordingly also have

$$\int \sum_{\mu,\nu} [\sqrt{g}J]_{\mu\nu} p^{\mu\nu} d\omega = 0.$$

Introduction of i_s, i_s^l into this identity finally shows that

$$\int \left(\sum_l \frac{\partial i_s^l}{\partial w_l} - i_s \right) p^s d\omega = 0$$

and therefore also that the assertion of our theorem is correct.

The most important aim is now the formulation of the concept of energy, and the derivation of the energy theorem solely on the basis of the two axioms I and II.

For this purpose we first form:

$$P_g(\sqrt{g}H) = \sum_{\mu,\nu,k,l} \left(\frac{\partial \sqrt{g}H}{\partial g^{\mu\nu}} p^{\mu\nu} + \frac{\partial \sqrt{g}H}{\partial g_k^{\mu\nu}} p_k^{\mu\nu} + \frac{\partial \sqrt{g}H}{\partial g_{kl}^{\mu\nu}} p_{kl}^{\mu\nu} \right).$$

Now $\frac{\partial H}{\partial g_{kl}^{\mu\nu}}$ is a mixed tensor of fourth rank, so if one puts

$$A_k^{\mu\nu} = p_k^{\mu\nu} + \sum_\rho \left(\begin{Bmatrix} k\rho \\ \mu \end{Bmatrix} p^{\rho\nu} + \begin{Bmatrix} k\rho \\ \nu \end{Bmatrix} p^{\rho\mu} \right),$$

$$\begin{Bmatrix} k\rho \\ \mu \end{Bmatrix} = \frac{1}{2} \sum_\sigma g^{\mu\sigma} (g_{k\sigma\rho} + g_{\rho\sigma k} - g_{k\rho\sigma}),$$

the expression

$$a^l = \sum_{\mu,\nu,k} \frac{\partial H}{\partial g_{kl}^{\mu\nu}} A_k^{\mu\nu} \tag{8}$$

becomes a contragredient vector.

Hence if we form the expression

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g}a^l}{\partial w_l}$$

[401] then this no longer contains the second derivatives $p_{kl}^{\mu\nu}$ and | therefore has the form

$$\sqrt{g} \sum_{\mu,\nu,k} (B_{\mu\nu} p^{\mu\nu} + B_{\mu\nu}^k p_k^{\mu\nu}),$$

where

$$B^k_{\mu\nu} = \sum_{\rho, l}\left(\frac{\partial H}{\partial g^{\mu\nu}_k} - \frac{\partial}{\partial w_l}\frac{\partial H}{\partial g^{\mu\nu}_{kl}} - \frac{\partial H}{\partial g^{\rho\nu}_{kl}}\left\{\begin{matrix} l\mu \\ \rho \end{matrix}\right\} - \frac{\partial H}{\partial g^{\mu\rho}_{kl}}\left\{\begin{matrix} l\nu \\ \rho \end{matrix}\right\}\right)$$

is again a mixed tensor.

Now we form the vector

$$b^l = \sum_{\mu, \nu} B^l_{\mu\nu} p^{\mu\nu}, \tag{9}$$

and obtain from it

$$P_g(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g}(a^l + b^l)}{\partial w_l} = \sum_{\mu, \nu}[\sqrt{g}H]_{\mu\nu} p^{\mu\nu}. \tag{10}$$

On the other hand we form

$$P_q(\sqrt{g}H) = \sum_{k, l}\left(\frac{\partial \sqrt{g}H}{\partial q_k} p_k + \frac{\partial \sqrt{g}H}{\partial q_{kl}} p_{kl}\right);$$

then $\dfrac{\partial H}{\partial q_{kl}}$ is a tensor and the expression

$$c^l = \sum_k \frac{\partial H}{\partial q_{kl}} p_k \tag{11}$$

therefore represents a contragredient vector. Correspondingly, as above, we obtain

$$P_q(\sqrt{g}H) - \sum_l \frac{\partial \sqrt{g} c^l}{dw_l} = \sum_k [\sqrt{g}H]_k p_k. \tag{12}$$

Now we note the basic equations (4) and (5), and conclude by adding (10) and (12):

$$P(\sqrt{g}H) = \sum_l \frac{\partial \sqrt{g}(a^l + b^l + c^l)}{dw_l}.$$

But we have

$$\begin{aligned} P(\sqrt{g}H) &= \sqrt{g}PH + H\sum_{\mu, \nu}\frac{\partial \sqrt{g}}{\partial g^{\mu\nu}} p^{\mu\nu} \\ &= \sqrt{g}PH + H\sum_s\left(\frac{\partial \sqrt{g}}{\partial w_s} p^s + \sqrt{g}\, p^s_s\right), \end{aligned}$$

and thus, due to identity (6)

$$P(\sqrt{g}H) = \sqrt{g}\sum_s \frac{\partial H}{\partial w_s}p^s + H\sum_s\left(\frac{\partial\sqrt{g}}{\partial w_s}p^s + \sqrt{g}\,p^s_s\right) = \sum_s \frac{\partial\sqrt{g}Hp^s}{\partial w_s}.$$

[402] | From this we finally obtain the invariant equation

$$\sum_l \frac{\partial}{\partial w_l}\sqrt{g}(Hp^l - a^l - b^l - c^l) = 0.$$

Now we note that

$$\frac{\partial H}{\partial q_{lk}} - \frac{\partial H}{\partial q_{kl}}$$

is a skew symmetric contravariant tensor; consequently

$$d^l = \frac{1}{2\sqrt{g}}\sum_{k,s}\frac{\partial}{\partial w_k}\left\{\left(\frac{\partial\sqrt{g}H}{\partial q_{lk}} - \frac{\partial\sqrt{g}H}{\partial q_{kl}}\right)p^s q_s\right\} \tag{13}$$

becomes a contravariant vector, which evidently satisfies the identity

$$\sum_l \frac{\partial\sqrt{g}d^l}{\partial w_l} = 0.$$

Let us now define

$$e^l = Hp^l - a^l - b^l - c^l - d^l \tag{14}$$

as the *energy vector, then the energy vector is a contravariant vector, which moreover depends linearly on the arbitrarily chosen vector* p^s, *and satisfies identically for that choice of this vector* p^s *the invariant energy equation*

$$\sum_l \frac{\partial\sqrt{g}e^l}{\partial w_l} = 0.$$

As far as the world function H is concerned, further axioms are needed to determine its choice in a unique way. If the gravitational field equations are to contain only second derivatives of the potentials $g^{\mu\nu}$, then H must have the form

$$H = K + L$$

where K is the invariant that derives from the Riemannian tensor (curvature of the four-dimensional manifold)

$$K = \sum_{\mu,\nu} g^{\mu\nu}K_{\mu\nu}$$

$$K_{\mu\nu} = \sum_{\kappa}\left(\frac{\partial}{\partial w_\nu}\left\{\begin{matrix}\mu\kappa\\ \kappa\end{matrix}\right\} - \frac{\partial}{\partial w_\kappa}\left\{\begin{matrix}\mu\nu\\ \kappa\end{matrix}\right\}\right) + \sum_{\kappa,\lambda}\left(\left\{\begin{matrix}\mu\kappa\\ \lambda\end{matrix}\right\}\left\{\begin{matrix}\lambda\nu\\ \kappa\end{matrix}\right\} - \left\{\begin{matrix}\mu\nu\\ \lambda\end{matrix}\right\}\left\{\begin{matrix}\lambda\kappa\\ \kappa\end{matrix}\right\}\right)$$

and where L depends only on $g^{\mu\nu}$, $g_l^{\mu\nu}$, q_s, q_{sk}. Finally we make the simplifying assumption in the following, that L does not contain the $g_l^{\mu\nu}$. |

Next we apply theorem II to the invariant L and obtain [403]

$$\sum_{\mu,\nu,m}\frac{\partial L}{\partial g^{\mu\nu}}(g^{\mu m}p_m^{\nu} + g^{\nu m}p_m^{\mu}) - \sum_{s,m}\frac{\partial L}{\partial q_s}q_m p_s^m - \sum_{s,k,m}\frac{\partial L}{\partial q_{sk}}(q_{sm}p_k^m + q_{mk}p_s^m + q_m p_{sk}^m) = 0. \tag{15}$$

Equating to zero the coefficient of p_{sk}^m on the left produces the equation

$$\left(\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}}\right)q_m = 0$$

or

$$\frac{\partial L}{\partial q_{sk}} + \frac{\partial L}{\partial q_{ks}} = 0, \tag{16}$$

that is, the derivatives of the electrodynamic potentials q_s occur only in the combinations

$$M_{ks} = q_{sk} - q_{ks}.$$

Thus we learn that under our assumptions the invariant L depends, besides on the potentials $g_{\mu\nu}$, q_s, only on the components of the skew symmetric invariant tensor

$$M = (M_{ks}) = \mathrm{Curl}(q_s),$$

that is, of the so-called electromagnetic six vector. *This result, which determines the character of Maxwell's equations in the first place, here derives essentially as a consequence of the general invariance, that is, on the basis of axiom II.*

If we put the coefficient of p_m^{ν} on the left of identity (15) equal to zero, we obtain, using (16)

$$2\sum_{\mu}\frac{\partial L}{\partial g^{\mu\nu}}g^{\mu m} - \frac{\partial L}{\partial q_m}q_\nu - \sum_s\frac{\partial L}{\partial M_{ms}}M_{\nu s} = 0, \qquad (\mu = 1, 2, 3, 4). \tag{17}$$

This equation admits an important transformation of the electromagnetic energy, that is the part of the energy vector that comes from L. Namely, this part results from (11), (13), (14) as follows:

$$Lp^l - \sum_k \frac{\partial L}{\partial q_{kl}} p_k - \frac{1}{2\sqrt{g}} \sum_{k,s} \frac{\partial}{\partial w_k} \left\{ \left(\frac{\partial \sqrt{g} L}{\partial q_{lk}} - \frac{\partial \sqrt{g} L}{\partial q_{kl}} \right) p^s q_s \right\}.$$

Because of (16) and by noting (5) this expression becomes |

[404]

$$\sum_{s,k} \left(L\delta_s^l - \frac{\partial L}{\partial M_{lk}} M_{sk} - \frac{\partial L}{\partial q_l} q_s \right) p^s \tag{18}$$

$$(\delta_s^l = 0, \quad l \neq s; \quad \delta_s^s = 1)$$

so because of (17) it equals

$$-\frac{2}{\sqrt{g}} \sum_{\mu,s} \frac{\partial \sqrt{g} L}{\partial g^{\mu s}} g^{\mu l} p^s. \tag{19}$$

Because of the formulas (21) to be developed below we see from this in particular that the electromagnetic energy, and therefore also the total energy vector e^l can be expressed through K alone, so that only the $g^{\mu\nu}$ and their derivatives, but not the q_s and their derivatives occur in it. If one takes the limit

$$g_{\mu\nu} = 0, \qquad (\mu \neq \nu)$$
$$g_{\mu\mu} = 1$$

in expression (18), then this limit agrees exactly with what Mie has proposed in his electrodynamics: *Mie's electromagnetic energy tensor is nothing but the generally invariant tensor that results from differentiation of the invariant L with respect to the gravitational potentials $g^{\mu\nu}$ in that limit*—a circumstance that gave me the first hint of the necessary close connection between Einstein's general relativity theory and Mie's electrodynamics, and which convinced me of the correctness of the theory here developed.

It remains to show directly how with the assumption

$$H = K + L \tag{20}$$

the generalized Maxwell equations (5) put forth above are entailed by the gravitational equations (4).

Using the notation introduced earlier for the variational derivatives with respect to the $g^{\mu\nu}$, the gravitational equations, because of (20), take the form

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial \sqrt{g} L}{\partial g^{\mu\nu}} = 0. \tag{21}$$

The first term on the left hand side becomes

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2} K g_{\mu\nu} \right),$$

| as follows easily without calculation from the fact that $K_{\mu\nu}$, apart from $g_{\mu\nu}$, is the [405]
only tensor of second rank and K the only invariant, that can be formed using only the $g^{\mu\nu}$ and their first and second differential quotients, $g^{\mu\nu}_k$, $g^{\mu\nu}_{kl}$.

The resulting differential equations of gravitation appear to me to be in agreement with the grand concept of the theory of general relativity established by Einstein in his later treatises.[5]

Further, if we denote in general the variational derivatives of $\sqrt{g}J$ with respect to the electrodynamic potential q_h as above by

$$[\sqrt{g}J]_h = \frac{\partial\sqrt{g}J}{\partial q_h} - \sum_k \frac{\partial}{\partial w_k}\frac{\partial\sqrt{g}J}{\partial q_{hk}},$$

then the basic electromagnetic equations assume the form, due to (20)

$$[\sqrt{g}L]_h = 0. \tag{22}$$

Since K is an invariant that depends only on the $g^{\mu\nu}$ and their derivatives, by theorem III the equation (7) holds identically, with

$$i_s = \sum_{\mu,\nu}[\sqrt{g}K]_{\mu\nu}g^{\mu\nu}_s \tag{23}$$

and

$$i^l_s = -2\sum_\mu [\sqrt{g}K]_{\mu s}g^{\mu l}, \qquad (\mu = 1, 2, 3, 4). \tag{24}$$

Due to (21) and (24), (19) equals $-\frac{1}{\sqrt{g}}i^m_\nu$. By differentiating with respect to w_m and summing over m we obtain because of (7)

$$\begin{aligned} i_\nu &= \sum_m \frac{\partial}{\partial w_m}\Big(-\sqrt{g}L\delta^m_\nu + \frac{\partial\sqrt{g}L}{\partial q_m}q_\nu + \sum_s \frac{\partial\sqrt{g}L}{\partial M_{sm}}M_{s\nu}\Big) \\ &= -\frac{\partial\sqrt{g}L}{\partial w_\nu} + \sum_m \Big\{ q_\nu \frac{\partial}{\partial w_m}\Big([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\Big) \\ &\quad + q_{\nu m}\Big([\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial\sqrt{g}L}{\partial q_{ms}}\Big)\Big\} \\ &\quad + \sum_s \Big([\sqrt{g}L]_s - \frac{\partial\sqrt{g}L}{\partial q_s}\Big)M_{s\nu} + \sum_{s,m}\frac{\partial\sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}, \end{aligned}$$

5 Loc. cit. *Berliner Sitzungsber.* 1915.

since of course

$$\frac{\partial \sqrt{g}L}{\partial q_m} = [\sqrt{g}L]_m + \sum_s \frac{\partial}{\partial w_s}\frac{\partial \sqrt{g}L}{\partial q_{ms}}$$

[406] | and[3]

$$-\sum_m \frac{\partial}{\partial w_m}\frac{\partial \sqrt{g}L}{\partial q_{sm}} = [\sqrt{g}L]_s - \frac{\partial \sqrt{g}L}{\partial q_s}.$$

Now we take into account that because of (16) we have

$$\sum_{m,\,s} \frac{\partial^2}{\partial w_m \partial w_s}\frac{\partial \sqrt{g}L}{\partial q_{ms}} = 0,$$

and then obtain by suitably collecting terms

$$\begin{aligned} i_\nu = &-\frac{\partial \sqrt{g}L}{\partial w_\nu} + \sum_m \left(q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m + M_{m\nu}[\sqrt{g}L]_m \right) \\ &+ \sum_m \frac{\partial \sqrt{g}L}{\partial q_m} q_{m\nu} + \sum_{s,\,m} \frac{\partial \sqrt{g}L}{\partial M_{sm}}\frac{\partial M_{s\nu}}{\partial w_m}. \end{aligned} \tag{25}$$

On the other hand we have

$$-\frac{\partial \sqrt{g}L}{\partial w_\nu} = -\sum_{s,\,m} \frac{\partial \sqrt{g}L}{\partial g^{sm}} g_\nu^{sm} - \sum_m \frac{\partial \sqrt{g}L}{\partial q_m} q_{m\nu} - \sum_{m,\,s} \frac{\partial \sqrt{g}L}{\partial q_{ms}}\frac{\partial q_{ms}}{\partial w_\nu}.$$

The first term on the right is nothing other than i_ν because of (21) and (23). The last term on the right proves to be equal and opposite to the last term on the right of (25); namely, we have

$$\sum_{s,\,m} \frac{\partial \sqrt{g}L}{\partial M_{sm}}\left(\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu} \right) = 0, \tag{26}$$

since the expression

$$\frac{\partial M_{s\nu}}{\partial w_m} - \frac{\partial q_{ms}}{\partial w_\nu} = \frac{\partial^2 q_\nu}{\partial w_s \partial w_m} - \frac{\partial^2 q_s}{\partial w_\nu \partial w_m} - \frac{\partial^2 q_m}{\partial w_\nu \partial w_s}$$

is symmetric in s, m, and the first factor under the summation sign in (26) turns out to be skew symmetric in s, m.

Consequently (25) entails the equation

$$\sum_m \left(M_{m\nu}[\sqrt{g}L]_m + q_\nu \frac{\partial}{\partial w_m}[\sqrt{g}L]_m \right) = 0; \tag{27}$$

that is, from the gravitational equations (4) there follow indeed the four mutually independent linear combinations (27) of the basic electrodynamic equations (5) and their first derivatives. *This is the exact mathematical expression of the statement claimed in general above concerning the character of electrodynamics as a consequence of gravitation.* |

According to our assumption L should not depend on the derivatives of the $g^{\mu\nu}$; [407] therefore L must be a function of certain four general invariants, which correspond to the special orthogonal invariants given by Mie, and of which the two simplest ones are these:

$$Q = \sum_{k,l,m,n} M_{mn} M_{lk} g^{mk} g^{nl}$$

and

$$q = \sum_{k,l} q_k q_l g^{kl}.$$

The simplest and most straightforward ansatz for L, considering the structure of K, is also that which corresponds to Mie's electrodynamics, namely

$$L = \alpha Q + f(q)$$

or, following Mie even more closely:

$$L = \alpha Q + \beta q^3,$$

where $f(q)$ denotes any function of q, and α, β are constants.

As one can see, the few simple assumptions expressed in axioms I and II suffice with appropriate interpretation to establish the theory: through it not only are our views of space, time, and motion fundamentally reshaped in the sense explained by Einstein, but I am also convinced that through the basic equations established here the most intimate, presently hidden processes in the interior of the atom will receive an explanation, and in particular that generally a reduction of all physical constants to mathematical constants must be possible—even as in the overall view thereby the possibility approaches that physics in principle becomes a science of the type of geometry: surely the highest glory of the axiomatic method, which as we have seen takes the powerful instruments of analysis, namely variational calculus and theory of invariants, into its service.

EDITORIAL NOTES

[1] The index $_l$ of $\partial\omega_l$ in the denominator of the third equation is missing in the original text.

[2] The subscript $_{sk}$ in the denominator of ∂q_{sk} is missing in the original text.

[3] The subscript $_s$ in the term $[\sqrt{g}L]_s$ is missing in the original text.

DAVID HILBERT

THE FOUNDATIONS OF PHYSICS (SECOND COMMUNICATION)

Originally published as "Die Grundlagen der Physik. (Zweite Mitteilung)" in Nachrichten von der Königlichen Gesellschaft zu Göttingen. Math.-phys. Klasse. 1917, p. 53–76. Presented in the Session of 23 December 1916.

In my first communication[1] I proposed a system of basic equations of physics. Before turning to the theory of integrating these equations it seems necessary to discuss some more general questions of a logical as well as physical nature.

First we introduce in place of the world parameters w_s $(s = 1, 2, 3, 4)$ the most general *real* spacetime coordinates x_s $(s = 1, 2, 3, 4)$ by putting

$$w_1 = x_1, \qquad w_2 = x_2, \qquad w_3 = x_3, \qquad w_4 = x_4,$$

and correspondingly in place of

$$ig_{14}, \qquad ig_{24}, \qquad ig_{34}, \qquad -g_{44},$$

we write simply

$$g_{14}, \qquad g_{24}, \qquad g_{34}, \qquad g_{44}.$$

The new $g_{\mu\nu}$ $(\mu, \nu = 1, 2, 3, 4)$ —the gravitational potentials of Einstein—shall then all be real functions of the real variables x_s $(s = 1, 2, 3, 4)$ of such a type that, in the representation of the quadratic form

$$G(X_1, X_2, X_3, X_4) = \sum_{\mu\nu} g_{\mu\nu} X_\mu X_\nu \tag{28}$$

as a sum of four squares of linear forms of the X_s, three squares always occur with positive sign, and one square with negative | sign: thus the quadratic form (28) pro- [54]
vides our four dimensional world of the x_s with the metric of a pseudo-geometry. The determinant g of the $g_{\mu\nu}$ turns out to be negative.

1 This journal, 20 November 1915.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

If a curve

$$x_s = x_s(p) \qquad (s = 1, 2, 3, 4)$$

is given in this geometry, where $x_s(p)$ mean some arbitrary real functions of the parameter p, then it can be divided into pieces of curves on each of which the expression

$$G(\frac{dx_1}{dp}, \frac{dx_2}{dp}, \frac{dx_3}{dp}, \frac{dx_4}{dp})$$

does not change sign: A piece of the curve for which

$$G(\frac{dx_s}{dp}) > 0$$

shall be called a *segment* and the integral along this piece of curve

$$\lambda = \int \sqrt{G(\frac{dx_s}{dp})}\, dp$$

shall be the *length of the segment*; a piece of the curve for which

$$G(\frac{dx_s}{dp}) < 0$$

will be called a *time line*, and the integral

$$\tau = \int \sqrt{-G(\frac{dx_s}{dp})}\, dp$$

evaluated along this piece of curve shall be the *proper time of the time line*; finally a piece of curve along which

$$G(\frac{dx_s}{dp}) = 0$$

shall be called a *null line*.

To visualize these concepts of our pseudo geometry we imagine two ideal measuring devices: the *measuring thread* by means of which we are able to measure the length λ of any segment, and secondly the *light clock* with which we can determine the proper time of any time line. The thread shows zero and the light clock stops along every null line, whereas the former fails totally along a time line, and the latter along a segment. |

[55] First we show that each of the two instruments suffices to compute with its aid the values of the $g_{\mu\nu}$ as functions of x_s, as soon as a definite spacetime coordinate system x_s has been introduced. Indeed we choose any set of 10 segments, which all converge on the same world point x_s, from different directions, so that this endpoint

assumes the same parameter value p on each. At this end point we have the equation, for each of the 10 segments,

$$\left(\frac{d\lambda^{(h)}}{dp}\right)^2 = G\left(\frac{dx_s^{(h)}}{dp}\right), \qquad (h = 1, 2, \ldots, 10);$$

here the left-hand sides are known as soon as we have determined the lengths $\lambda^{(h)}$ by means of the thread. We introduce the abbreviations

$$D(u) = \begin{vmatrix} \left(\frac{dx_1^{(1)}}{dp}\right)^2, & \frac{dx_1^{(1)}}{dp}\frac{dx_2^{(1)}}{dp}, & \ldots, & \left(\frac{dx_4^{(1)}}{dp}\right)^2, & \left(\frac{d\lambda^{(1)}}{dp}\right)^2 \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots & \ldots\ldots \\ \left(\frac{dx_1^{(10)}}{dp}\right)^2, & \frac{dx_1^{(10)}}{dp}\frac{dx_2^{(10)}}{dp}, & \ldots, & \left(\frac{dx_4^{(10)}}{dp}\right)^2, & \left(\frac{d\lambda^{(10)}}{dp}\right)^2 \\ X_1^2, & X_1X_2, & \ldots, & X_4^2, & u \end{vmatrix},$$

so that clearly

$$G(X_s) = -\frac{D(0)}{\frac{\partial D}{\partial u}}, \tag{29}$$

whereby also the condition on the directions of the chosen 10 segments at the point $x_s(p)$

$$\frac{\partial D}{\partial u} \neq 0$$

is seen to be necessary.

When G has been calculated according to (29), the use of this procedure for any 11th segment ending at $x_s(p)$ would yield the equation

$$\left(\frac{d\lambda^{(11)}}{dp}\right)^2 = G\left(\frac{dx_s^{(11)}}{dp}\right),$$

and this equation would then both verify the correctness of the instrument and confirm experimentally that the postulates of the theory apply to the real world.

Corresponding reasoning applies to the light clock. |

[56] The axiomatic construction of our pseudo-geometry could be carried out without difficulty: first an axiom should be established from which it follows that length resp. proper time must be integrals whose integrand is only a function of the x_s and their first derivatives with respect to the parameter; suitable for such an axiom would be the property of development of the thread or the well-known envelope theorem for geodesic lines. Secondly an axiom is needed whereby the theorems of the pseudo-Euclidean geometry, that is the old principle of relativity, shall be valid in infinitesi-

mal regions; for this the axiom put down by W. Blaschke[2] would be particularly suitable, which states that the condition of orthogonality for any two directions—segments or time lines—shall always be a symmetric relation.

Let us briefly summarize the main facts that the Monge-Hamilton theory of differential equations teaches us for our pseudo-geometry.

With every world point x_s there is associated a cone of second order, with vertex at x_s, and determined in the running point coordinates X_s by the equation

$$G(X_1 - x_1, X_2 - x_2, X_3 - x_3, X_4 - x_4) = 0;$$

this shall be called the *null cone* belonging to the point x_s. The totality of null cones form a four dimensional field of cones, which is associated on the one hand with "Monge's" differential equation

$$G(\frac{dx_1}{dp}, \frac{dx_2}{dp}, \frac{dx_3}{dp}, \frac{dx_4}{dp}) = 0,$$

and on the other hand with "Hamilton's" partial differential equation

$$H(\frac{df}{dx_1}, \frac{df}{dx_2}, \frac{df}{dx_3}, \frac{df}{dx_4}) = 0, \tag{30}$$

where H denotes the quadratic form

$$H(U_1, U_2, U_3, U_4) = \sum_{\mu\nu} g^{\mu\nu} U_\mu U_\nu$$

reciprocal to G. The characteristics of Monge's and at the same time those of Hamilton's partial differential equation (30) are the geodesic null lines. All the geodesic null lines originating at one particular world point a_s $(s = 1, 2, 3, 4)$ generate a three
[57] dimensional point manifold, which I shall be called the *time divide* belonging to the world point a_s. This divide has a node at a_s, whose tangent cone is precisely the null cone belonging to a_s. If we transform the equation of the time divide into the form

$$x_4 = \varphi(x_1, x_2, x_3),$$

then

$$f = x_4 - \varphi(x_1, x_2, x_3)$$

is an integral of Hamilton's differential equation (30). All the time lines originating at the point a_s remain totally in the interior of that four dimensional part of the world whose boundary is the time divide of a_s.

After these preparations we turn to the problem of *causality* in the new physics.

---

2 "Räumliche Variationsprobleme mit symmetrischer Transversalitätsbedingung." *Leipziger Berichte, Math.-phys. Kl.* 68 (1916) p. 50.

Up to now all coordinate systems x_s, that result from any one by arbitrary transformation have been regarded as equally valid. This arbitrariness must be restricted when we want to realize the concept that two world points on the same time line can be related as cause and effect, and that it should then no longer be possible to transform such world points to be simultaneous. In declaring x_4 as the *true* time coordinate we adopt the following definition:

A *true* spacetime coordinate system is one for which the following four inequalities hold, in addition to $g < 0$:

$$g_{11} > 0, \qquad \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} > 0, \qquad \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} > 0, \qquad g_{44} < 0. \tag{31}$$

A transformation that transforms one such spacetime coordinate system into another true spacetime coordinate system shall be called a *true* spacetime coordinate transformation.

The four inequalities mean that at any world point a_s the associated null cone excludes the linear space

$$x_4 = a_4,$$

but contains in its interior the line

$$x_1 = a_1, \qquad x_2 = a_2, \qquad x_3 = a_3;$$

the latter line is therefore always a time line. |

Let any time line $x_s = x_s(p)$ be given; because [58]

$$G\left(\frac{dx_s}{dp}\right) < 0$$

it follows that in a true spacetime coordinate system we must always have

$$\frac{dx_4}{dp} \neq 0,$$

and therefore that along a time line the true time coordinate x_4 must always increase resp. decrease. Because a time line remains a time line upon every coordinate transformation, therefore two world points along one time line can never be given the same value of the time coordinate x_4 through a true spacetime transformation; that is, they cannot be transformed to be simultaneous.

On the other hand, if the points of a curve can be truly transformed to be simultaneous, then after this transformation we have for this curve

$$x_4 = \text{const.}, \qquad \text{that is} \qquad \frac{dx_4}{dp} = 0,$$

therefore

$$G\left(\frac{dx_s}{dp}\right) = \sum_{\mu\nu} g_{\mu\nu}\frac{dx_\mu}{dp}\frac{dx_\nu}{dp}, \qquad (\mu, \nu = 1, 2, 3),$$

and here the right side is positive because of the first three of our inequalities (31); the curve therefore characterizes a *segment*.

So we see that the concepts of cause and effect, which underlie the principle of causality, also do not lead to any inner contradictions whatever in the new physics, if we only take the inequalities (31) always to be part of our basic equations, that is if we confine ourselves to using *true* spacetime coordinates.

At this point let us take note of a special spacetime coordinate system that will later be useful and which I will call the *Gaussian coordinate system*, because it is the generalization of the system of geodesic polar coordinates introduced by Gauss in the theory of surfaces. In our four-dimensional world let any three-dimensional space be given so that every curve confined to that space is a segment: *a space of segments*, as
[59] I would like to call it; | let x_1, x_2, x_3 be any point coordinates of this space. We now construct at every point x_1, x_2, x_3 of this space the geodesic orthogonal to it, which will be a time line, and on this line we mark off x_4 as proper time; the point in the four-dimensional world so obtained is given coordinate values $x_1x_2x_3x_4$. In these coordinates we have, as is easily seen,

$$G(X_s) = \sum_{\mu\nu}^{1,2,3} g_{\mu\nu}X_\mu X_\nu - X_4^2, \tag{32}$$

that is, the Gaussian coordinate system is characterized analytically by the equations

$$g_{14} = 0, \qquad g_{24} = 0, \qquad g_{34} = 0, \qquad g_{44} = 0. \tag{33}$$

Because of the nature of the three dimensional space $x_4 = 0$ we presupposed, the quadratic form on the right-hand side of (32) in the variables X_1, X_2, X_3 is necessarily positive definite, so the first three of the inequalities (31) are satisfied, and since this also applies to the fourth, the Gaussian coordinate system always turns out to be a *true* spacetime coordinate system.

We now return to the investigation of the principle of causality in physics. As its main contents we consider the fact, valid so far in every physical theory, that from a knowledge of the physical quantities and their time derivatives in the present the future values of these quantities can always be determined: without exception the laws of physics to date have been expressed in a system of differential equations in which the number of the functions occurring in them was essentially the same as the number of independent differential equations; and thus the well-known general Cauchy theorem on the existence of integrals of partial differential equations directly offered the rationale of proof for the above fact.

Now, as I emphasized particularly in my first communication, the basic equations of physics (4) and (5) established there are by no means of the type characterized

above; rather, according to Theorem I, four of them are a consequence of the rest: we regarded the four Maxwell equations (5) as a consequence of the ten gravitational equations (4), and so we have for the 14 potentials $g_{\mu\nu}$, q_s only 10 equations (4) that are essentially independent of each other. |

As soon as we maintain the demand of general invariance for the basic equations [60]
of physics the circumstance just mentioned is essential and even necessary. Because if there were further invariant equations, independent of (4), for the 14 potentials, then introduction of a Gaussian coordinate system would lead for the 10 physical quantities as per (33),

$$g_{\mu\nu} \quad (\mu, \nu = 1, 2, 3), \qquad q_s \quad (s = 1, 2, 3, 4)$$

to a system of equations that would again be mutually independent, and mutually contradictory, because there are more than 10 of them.

Under such circumstances then, as occur in the new physics of general relativity, it is by no means any longer possible from knowledge of physical quantities in present and past to derive uniquely their future values. To show this intuitively on an example, let our basic equations (4) and (5) of the first communication be integrated in the special case corresponding to the presence of a single electron permanently at rest, so that the 14 potentials

$$g_{\mu\nu} = g_{\mu\nu}(x_1, x_2, x_3)$$
$$q_s = q_s(x_1, x_2, x_3)$$

become definite functions of x_1, x_2, x_3, all independent of the time x_4, and in addition such that the first three components r_1, r_2, r_3 of the four-current density vanish. Then we apply the following coordinate transformation to these potentials:

$$\begin{cases} x_1 = x'_1 & \text{for } x'_4 \le 0 \\ x_1 = x'_1 + e^{-\frac{1}{x'^2_4}} & \text{for } x'_4 > 0 \end{cases}$$
$$x_2 = x'_2$$
$$x_3 = x'_3$$
$$x_4 = x'_4.$$

For $x'_4 \le 0$ the transformed potentials $g'_{\mu\nu}$, q'_s are the same functions of x'_1, x'_2, x'_3 as the $g_{\mu\nu}$, q_s of the original variables x_1, x_2, x_3, whereas the $g'_{\mu\nu}$, q'_s for $x'_4 > 0$ depend in an essential way also on the time coordinate x'_4; that is, the potentials $g'_{\mu\nu}$, q'_s represent an electron that is at rest until $x'_4 = 0$, but then puts its components into motion. |

Nonetheless I believe that it is only necessary to formulate more sharply the idea [61]
on which the principle of general relativity[3] is based, in order to maintain the principle of causality also in the new physics. Namely, to follow the essence of the new rel-

ativity principle we must demand invariance not only for the general laws of physics, but we must accord invariance to each separate statement in physics that is to have physical meaning—in accordance with this, that in the final analysis it must be possible to establish each physical fact by thread or light clock, that is, instruments of *invariant* character. In the theory of curves and surfaces, where a statement in a chosen parametrization of the curve or surface has no geometrical meaning for the curve or surface itself, if this statement does not remain invariant under any arbitrary transformation of the parameters or cannot be brought to invariant form; so also in physics we must characterize a statement that does not remain invariant under any arbitrary transformation of the coordinate system as *physically meaningless*. For example, in the case considered above of the electron at rest, the statement that, say at the time $x_4 = 1$ this electron is at rest, has no physical meaning because this statement is not invariant.

Concerning the principle of causality, let the physical quantities and their time derivatives be known at the present in some given coordinate system: then a statement will only have physical meaning if it is invariant under all those transformations, for which the coordinates just used for the present remain unchanged; I maintain that statements of this type for the future are all uniquely determined, that is, *the principle of causality holds in this form*:

From present knowledge of the 14 physical potentials $g_{\mu\nu}$, q_s *all statements about them for the future follow necessarily and uniquely provided they are physically meaningful.*

To prove this proposition we use the *Gaussian* spacetime coordinate system. Introducing (33) into the basic equations (4) of the first communication yields for the 10 potentials |

[62] $$g_{\mu\nu} \quad (\mu, \nu = 1, 2, 3), \qquad q_s \quad (s = 1, 2, 3, 4) \tag{34}$$

a system of as many partial differential equations; if we integrate these on the basis of the given initial values at $x_4 = 0$, we find uniquely the values of (34) for $x_4 > 0$. Since the Gaussian coordinate system itself is uniquely determined, therefore also all statements about those potentials (34) with respect to these coordinates are of invariant character.

The forms, in which physically meaningful, i.e. invariant, statements can be expressed mathematically are of great variety.

First. This can be done by means of an invariant coordinate system. Like the Gaussian system used above one can apply the well-known Riemannian one, as well as that spacetime coordinate system in which electricity appears at rest with unit current density. As at the end of the first communication, let $f(q)$ denote the function occurring in Hamilton's principle and depending on the invariant

---

3 In his original theory, now abandoned, A. Einstein (*Sitzungsberichte der Akad. zu Berlin*, 1914, p. 1067) had indeed postulated certain 4 non-invariant equations for the $g_{\mu\nu}$, in order to save the causality principle in its old form.

$$q = \sum_{kl} q_k q_l g^{kl},$$

then

$$r^s = \frac{\partial f(q)}{\partial q_s}$$

is the four-current density of electricity; it represents a contravariant vector and therefore can certainly be transformed to $(0, 0, 0, 1)$, as is easily seen. If this is done, then from the four equations

$$\frac{\partial f(q)}{\partial q_s} = 0 \quad (s = 1, 2, 3), \qquad \frac{\partial f(q)}{\partial q_4} = 1$$

the four components of the four-potential q_s can be expressed in terms of the $g_{\mu\nu}$, and every relation between the $g_{\mu\nu}$ in this or in one of the first two coordinate systems is then an invariant statement. For particular solutions of the basic equations there may be special invariant coordinate systems; for example, in the case treated below of the centrally symmetric gravitational field r, ϑ, φ, t form an invariant system of coordinates up to rotations.

Second. The statement, *according to which a coordinate system can be found* in which the 14 potentials $g_{\mu\nu}$, q_s have certain definite values in the future, or fulfill certain definite conditions, is always an invariant and therefore a physically meaningful one. The mathematically invariant expression for | such a statement is obtained by [63]
eliminating the coordinates from those relations. The case considered above, of the electron at rest, provides an example: the essential and physically meaningful content of the causality principle is here expressed by the statement that the electron which is at rest for the time $x_4 \leq 0$ will, *for a suitably chosen spacetime coordinate system*, also remain at rest in all its parts for the future $x_4 > 0$.

Third. A statement is also invariant and thus has physical meaning if it is supposed to be valid in any arbitrary coordinate system. An example of this are Einstein's energy-momentum equations having divergence character. For, although Einstein's energy does not have the property of invariance, and the differential equations he put down for its components are by no means covariant as a system of equations, nevertheless the assertion contained in them, that they shall be satisfied in any coordinate system, is an invariant demand and therefore it carries physical meaning.

According to my exposition, physics is a four-dimensional pseudo-geometry, whose metric $g_{\mu\nu}$ is connected to the electromagnetic quantities, i.e. to the matter, by the basic equations (4) and (5) of my first communication. With this understanding, an old geometrical question becomes ripe for solution, namely whether and in what sense Euclidean geometry—about which we know from mathematics only that it is a logical structure free from contradictions—also possesses validity in the real world.

The old physics with the concept of absolute time took over the theorems of Euclidean geometry and without question put them at the basis of every physical theory. Gauss as well proceeded hardly differently: he constructed a hypothetical non-

Euclidean physics, by maintaining the absolute time and revoking only the parallel axiom from the propositions of Euclidean geometry; a measurement of the angles of a triangle of large dimensions showed him the invalidity of this non-Euclidean physics.

The new physics of Einstein's principle of general relativity takes a totally different position vis-à-vis geometry. It takes neither Euclid's nor any other particular geometry *a priori* as basic, in order to deduce from it the proper laws of physics, but,
[64] as I showed in my first communication, | the new physics provides at one fell swoop through one and the same Hamilton's principle the geometrical and the physical laws, namely the basic equations (4) and (5), which tell us how the metric $g_{\mu\nu}$ —at the same time the mathematical expression of the phenomenon of gravitation—is connected with the values q_s of the electrodynamic potentials.

Euclidean geometry is *an action-at-a-distance law foreign to the modern physics*: By revoking the Euclidean geometry as a general presupposition of physics, the theory of relativity maintains instead that geometry and physics have identical character and are based as *one* science on a common foundation.

The geometrical question mentioned above amounts to the investigation, whether and under what conditions the four-dimensional Euclidean pseudo-geometry

$$\begin{aligned} &g_{11} = 1, \qquad g_{22} = 1, \qquad g_{33} = 1, \qquad g_{44} = -1 \\ &g_{\mu\nu} = 0 \qquad (\mu \neq \nu) \end{aligned} \tag{35}$$

is a solution, or even the only regular solution, of the basic physical equations.

The basic equations (4) of my first communication are, due to the assumption (20) made there:

$$[\sqrt{g}K]_{\mu\nu} + \frac{\partial \sqrt{g}L}{\partial g^{\mu\nu}} = 0,$$

where

$$[\sqrt{g}K]_{\mu\nu} = \sqrt{g}\left(K_{\mu\nu} - \frac{1}{2}Kg_{\mu\nu}\right).$$

When the values (35) are substituted, we have

$$[\sqrt{g}K]_{\mu\nu} = 0 \tag{36}$$

and for

$$q_s = 0 \qquad (s = 1, 2, 3, 4)$$

we have

$$\frac{\partial \sqrt{g}L}{\partial g^{\mu\nu}} = 0;$$

that is, when all electricity is removed, the pseudo-Euclidean geometry is possible. The question whether it is also necessary in this case, i.e. whether—or under certain

additional conditions—the values (35), and those values of the $g_{\mu\nu}$ resulting from coordinate transformation of the latter, are the only regular solutions of the equations (36) is a mathematical problem not to be discussed here in general. Instead I confine myself | to presenting some thoughts concerning this problem in particular. [65]

For this we return to the original world coordinates of my first communication

$$w_1 = x_1, \qquad w_2 = x_2, \qquad w_3 = x_3, \qquad w_4 = ix_4,$$

and give the corresponding meaning to the $g_{\mu\nu}$.

In the case of the pseudo-Euclidean geometry we have

$$g_{\mu\nu} = \delta_{\mu\nu},$$

where

$$\delta_{\mu\nu} = 1, \qquad \delta_{\mu\nu} = 0 \quad (\mu \neq \nu).$$

For every metric in the neighborhood of this pseudo-Euclidean geometry the ansatz

$$g_{\mu\nu} = \delta_{\mu\nu} + \varepsilon h_{\mu\nu} + \ldots \tag{37}$$

is valid, where ε is a quantity converging to zero, and $h_{\mu\nu}$ are functions of the w_s. I make the following two assumptions about the metric (37):

I. The $h_{\mu\nu}$ shall be independent of the variable w_4.

II. The $h_{\mu\nu}$ shall show a certain regular behavior at infinity.

Now, if the metric (37) is to satisfy the differential equation (36) for all ε then it follows that the $h_{\mu\nu}$ must necessarily satisfy certain linear homogeneous partial differential equations of second order. If we substitute, following Einstein[4]

$$h_{\mu\nu} = k_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}\sum_s k_{ss}, \qquad (k_{\mu\nu} = k_{\nu\mu}) \tag{38}$$

and assume among the 10 functions $k_{\mu\nu}$ the four relations

$$\sum_s \frac{dk_{\mu s}}{dw_s} = 0, \qquad (\mu = 1, 2, 3, 4) \tag{39}$$

then these differential equations become:

$$\Box k_{\mu\nu} = 0, \tag{40}$$

where the abbreviation

4 "Näherungsweise Integration der Feldgleichungen der Gravitation." *Berichte d. Akad. zu Berlin* 1916, p. 688.

$$\square = \sum_s \frac{\partial^2}{\partial w_s^2}$$

has been used.

Because of the ansatz (38) the relations (39) are restrictive assumptions for the
[66] functions $h_{\mu\nu}$; however I will | show how one can always achieve, by suitable infinitesimal transformation of the variables w_1, w_2, w_3, w_4, that those restrictive assumptions are satisfied for the corresponding functions $h'_{\mu\nu}$ after the transformation.

To this end one should determine four functions $\varphi_1, \varphi_2, \varphi_3, \varphi_4$, which satisfy respectively the differential equations

$$\square \varphi_\mu = \frac{1}{2}\frac{\partial}{\partial w_\mu}\sum_\nu h_{\nu\nu} - \sum_\nu \frac{\partial h_{\mu\nu}}{\partial w_\nu}. \tag{41}$$

By means of the infinitesimal transformation

$$w_s = w'_s + \varepsilon\varphi_s,$$

$g_{\mu\nu}$ becomes

$$g'_{\mu\nu} = g_{\mu\nu} + \varepsilon\sum_\alpha g_{\alpha\nu}\frac{\partial\varphi_\alpha}{\partial w_\mu} + \varepsilon\sum_\alpha g_{\alpha\mu}\frac{\partial\varphi_\alpha}{\partial w_\nu} + \dots$$

or because of (37) it becomes

$$g'_{\mu\nu} = \delta_{\mu\nu} + \varepsilon h'_{\mu\nu} + \dots,$$

where I have put

$$h'_{\mu\nu} = h_{\mu\nu} + \frac{\partial\varphi_\nu}{\partial w_\mu} + \frac{\partial\varphi_\mu}{\partial w_\nu}.$$

If we now choose

$$k_{\mu\nu} = h'_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}\sum_s h'_{ss},$$

then these functions satisfy Einstein's condition (39) because of (41), and we have

$$h'_{\mu\nu} = k_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}\sum_s k_{ss} \qquad (k_{\mu\nu} = k_{\nu\mu}).$$

The differential equations (40), which must be valid according to the above argument for the $k_{\mu\nu}$ we found, become due to assumption I

$$\frac{\partial^2 k_{\mu\nu}}{\partial w_1^2} + \frac{\partial^2 k_{\mu\nu}}{\partial w_2^2} + \frac{\partial^2 k_{\mu\nu}}{\partial w_3^2} = 0,$$

and, since assumption II—*mutatis mutandis*—allows the conclusion that the $k_{\mu\nu}$ approach constants at infinity, it follows that these must be constant in general, that is: *By varying the metric of the pseudo-Euclidean geometry under the assumptions I and II it is not possible to obtain a regular metric that is not likewise pseudo-Euclidean and which also corresponds to a world free of electricity.* |

The integration of the partial differential equations (36) can be performed in yet [67] another case, first treated by Einstein[5] and by Schwarzschild.[6] In the following I present for this case a procedure that makes no assumptions about the gravitational potentials $g_{\mu\nu}$ at infinity, and which moreover offers advantages for my later investigations. The assumptions about the $g_{\mu\nu}$ are the following:

1. The metric is represented in a Gaussian coordinate system, except that g_{44} is left arbitrary, i.e. we have

$$g_{14} = 0, \qquad g_{24} = 0, \qquad g_{34} = 0.$$

2. The $g_{\mu\nu}$ are independent of the time coordinate x_4.
3. The gravitation $g_{\mu\nu}$ is centrally symmetric with respect to the origin of coordinates.

According to Schwarzschild the most general metric conforming to these assumptions is represented in polar coordinates, where

$$\begin{aligned} w_1 &= r\cos\vartheta \\ w_2 &= r\sin\vartheta\cos\varphi \\ w_3 &= r\sin\vartheta\sin\varphi \\ w_4 &= l, \end{aligned}$$

by the expression

$$F(r)dr^2 + G(r)(d\vartheta^2 + \sin^2\vartheta d\varphi^2) + H(r)dl^2 \tag{42}$$

where $F(r)$, $G(r)$, $H(r)$ are still arbitrary functions of r. If we put

$$r^* = \sqrt{G(r)},$$

then we are equally justified in interpreting r^*, ϑ, φ as spatial polar coordinates. If we introduce r^* in (42) instead of r and then eliminate the sign $*$, the result is the expression

$$M(r)dr^2 + r^2 d\vartheta^2 + r^2\sin^2\vartheta d\varphi^2 + W(r)dl^2, \tag{43}$$

5 "Perihelbewegung des Merkur." *Situngsber. d. Akad. zu Berlin*. 1915, p. 831.

6 "Über das Gravitationsfeld eines Massenpunktes." *Sitzungsber. d. Akad. zu Berlin*. 1916, p. 189.

where $M(r)$, $W(r)$ mean the two essential, arbitrary functions of r. The question is whether and how these can be determined in the most general way so that the differential equations (36) enjoy satisfaction. |

[68] To this end the well-known expressions $K_{\mu\nu}$, K given in my first communication must be calculated. The first step in this is the derivation of the differential equations for geodesic lines by variation of the integral

$$\int\left(M\left(\frac{dr}{dp}\right)^2+r^2\left(\frac{d\vartheta}{dp}\right)^2+r^2\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2+W\left(\frac{dl}{dp}\right)^2\right)dp.$$

As Lagrange equations we obtain these:

$$\begin{aligned}
&\frac{d^2r}{dp^2}+\frac{1}{2}\frac{M'}{M}\left(\frac{dr}{dp}\right)^2-\frac{r}{M}\left(\frac{d\vartheta}{dp}\right)^2-\frac{r}{M}\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2-\frac{1}{2}\frac{W'}{M}\left(\frac{dl}{dp}\right)^2=0,\\
&\frac{d^2\vartheta}{dp^2}+\frac{2}{r}\frac{dr}{dp}\frac{d\vartheta}{dp}-\sin\vartheta\cos\vartheta\left(\frac{d\varphi}{dp}\right)^2=0,\\
&\frac{d^2\varphi}{dp^2}+\frac{2}{r}\frac{dr}{dp}\frac{d\varphi}{dp}+2\cot\vartheta\frac{d\vartheta}{dp}\frac{d\varphi}{dp}=0,\\
&\frac{d^2l}{dp^2}+\frac{W'}{W}\frac{dr}{dp}\frac{dl}{dp}=0;
\end{aligned}$$

here and in the following calculation the sign $'$ denotes the derivative with respect to r. By comparison with the general differential equations of geodesic lines:

$$\frac{d^2w_s}{dp^2}+\sum_{\mu\nu}\left\{\begin{matrix}\mu\nu\\ s\end{matrix}\right\}\frac{dw_\mu}{dp}\frac{dw_\nu}{dp}=0,$$

we obtain for the bracket symbols $\left\{\begin{matrix}\mu\nu\\ s\end{matrix}\right\}$ the following values, whereby those that vanish are omitted:

$$\left\{\begin{matrix}11\\1\end{matrix}\right\}=\frac{1}{2}\frac{M'}{M},\qquad\left\{\begin{matrix}22\\1\end{matrix}\right\}=-\frac{r}{M},\qquad\left\{\begin{matrix}33\\1\end{matrix}\right\}=-\frac{r}{M}\sin^2\vartheta,$$

$$\left\{\begin{matrix}44\\1\end{matrix}\right\}=-\frac{1}{2}\frac{W'}{M},\qquad\left\{\begin{matrix}12\\2\end{matrix}\right\}=\frac{1}{r},\qquad\left\{\begin{matrix}33\\2\end{matrix}\right\}=-\sin\vartheta\cos\vartheta,$$

$$\left\{\begin{matrix}13\\3\end{matrix}\right\}=\frac{1}{r},\qquad\left\{\begin{matrix}23\\3\end{matrix}\right\}=\cot\vartheta,\qquad\left\{\begin{matrix}14\\4\end{matrix}\right\}=\frac{1}{2}\frac{W'}{W}.$$

With these we form:

$$
\begin{aligned}
K_{11} &= \frac{\partial}{\partial r}\left(\left\{\begin{matrix}11\\1\end{matrix}\right\}+\left\{\begin{matrix}12\\2\end{matrix}\right\}+\left\{\begin{matrix}13\\3\end{matrix}\right\}+\left\{\begin{matrix}14\\4\end{matrix}\right\}\right)-\frac{\partial}{\partial r}\left\{\begin{matrix}11\\1\end{matrix}\right\}\\
&\quad+\left\{\begin{matrix}11\\1\end{matrix}\right\}\left\{\begin{matrix}11\\1\end{matrix}\right\}+\left\{\begin{matrix}12\\2\end{matrix}\right\}\left\{\begin{matrix}21\\2\end{matrix}\right\}+\left\{\begin{matrix}13\\3\end{matrix}\right\}\left\{\begin{matrix}31\\3\end{matrix}\right\}+\left\{\begin{matrix}14\\4\end{matrix}\right\}\left\{\begin{matrix}41\\4\end{matrix}\right\}\\
&\quad-\left\{\begin{matrix}11\\1\end{matrix}\right\}\left(\left\{\begin{matrix}11\\1\end{matrix}\right\}+\left\{\begin{matrix}12\\2\end{matrix}\right\}+\left\{\begin{matrix}13\\3\end{matrix}\right\}+\left\{\begin{matrix}14\\4\end{matrix}\right\}\right)\\
&= \frac{1}{2}\frac{W''}{W}+\frac{1}{4}\frac{W'^2}{W^2}-\frac{M'}{rM}-\frac{1}{4}\frac{M'W'}{MW}
\end{aligned}
$$

$$
\begin{aligned}
K_{22} &= \frac{\partial}{\partial \vartheta}\left\{\begin{matrix}23\\3\end{matrix}\right\}-\frac{\partial}{\partial r}\left\{\begin{matrix}22\\1\end{matrix}\right\}\\
&\quad+\left\{\begin{matrix}21\\2\end{matrix}\right\}\left\{\begin{matrix}22\\1\end{matrix}\right\}+\left\{\begin{matrix}22\\1\end{matrix}\right\}\left\{\begin{matrix}12\\2\end{matrix}\right\}+\left\{\begin{matrix}23\\3\end{matrix}\right\}\left\{\begin{matrix}32\\3\end{matrix}\right\}\\
&\quad-\left\{\begin{matrix}22\\1\end{matrix}\right\}\left(\left\{\begin{matrix}11\\1\end{matrix}\right\}+\left\{\begin{matrix}12\\2\end{matrix}\right\}+\left\{\begin{matrix}13\\3\end{matrix}\right\}+\left\{\begin{matrix}14\\4\end{matrix}\right\}\right)\\
&= -1+\frac{1}{2}\frac{rM'}{M^2}+\frac{1}{M}+\frac{1}{2}\frac{rW'}{MW}
\end{aligned}
\qquad [69]
$$

$$
\begin{aligned}
K_{33} &= -\frac{\partial}{\partial r}\left\{\begin{matrix}33\\1\end{matrix}\right\}-\frac{\partial}{\partial \vartheta}\left\{\begin{matrix}33\\2\end{matrix}\right\}\\
&\quad+\left\{\begin{matrix}31\\3\end{matrix}\right\}\left\{\begin{matrix}33\\1\end{matrix}\right\}+\left\{\begin{matrix}32\\3\end{matrix}\right\}\left\{\begin{matrix}33\\2\end{matrix}\right\}+\left\{\begin{matrix}33\\1\end{matrix}\right\}\left\{\begin{matrix}13\\3\end{matrix}\right\}+\left\{\begin{matrix}33\\2\end{matrix}\right\}\left\{\begin{matrix}23\\3\end{matrix}\right\}\\
&\quad-\left\{\begin{matrix}33\\1\end{matrix}\right\}\left(\left\{\begin{matrix}11\\1\end{matrix}\right\}+\left\{\begin{matrix}12\\2\end{matrix}\right\}+\left\{\begin{matrix}13\\3\end{matrix}\right\}+\left\{\begin{matrix}14\\4\end{matrix}\right\}\right)-\left\{\begin{matrix}33\\2\end{matrix}\right\}\left\{\begin{matrix}23\\3\end{matrix}\right\}\\
&= \sin^2\vartheta\left(-1-\frac{1}{2}\frac{rM'}{M^2}+\frac{1}{M}+\frac{1}{2}\frac{rW'}{MW}\right)
\end{aligned}
$$

$$K_{44} = -\frac{\partial}{\partial r}\left\{\begin{matrix}44\\1\end{matrix}\right\} + \left\{\begin{matrix}41\\4\end{matrix}\right\}\left\{\begin{matrix}44\\1\end{matrix}\right\} + \left\{\begin{matrix}44\\1\end{matrix}\right\}\left\{\begin{matrix}41\\4\end{matrix}\right\}$$
$$- \left\{\begin{matrix}44\\1\end{matrix}\right\}\left(\left\{\begin{matrix}11\\1\end{matrix}\right\} + \left\{\begin{matrix}12\\2\end{matrix}\right\} + \left\{\begin{matrix}13\\3\end{matrix}\right\} + \left\{\begin{matrix}14\\4\end{matrix}\right\}\right)$$
$$= \frac{1}{2}\frac{W''}{M} - \frac{1}{4}\frac{M'W'}{M^2} - \frac{1}{4}\frac{W'^2}{MW} + \frac{W'}{rM}$$

$$K = \sum_s g^{ss}K_{ss} = \frac{W''}{MW} - \frac{1}{2}\frac{W'^2}{MW^2} - 2\frac{M'}{rM^2} - \frac{1}{2}\frac{M'W'}{M^2W}$$
$$-\frac{2}{r^2} + \frac{2}{r^2M} + 2\frac{W'}{rMW}.$$

Because

$$\sqrt{g} = \sqrt{MW}\,r^2\sin\vartheta$$

we have

$$K\sqrt{g} = \left\{\left(\frac{r^2W'}{\sqrt{MW}}\right)' - 2\frac{rM'\sqrt{W}}{M^{3/2}} - 2\sqrt{MW} + 2\sqrt{\frac{W}{M}}\right\}\sin\vartheta,$$

and if we put

$$M = \frac{r}{r-m}, \qquad W = w^2\frac{r-m}{r},$$

where now m and w are the unknown functions of r, we finally obtain

$$K\sqrt{g} = \left\{\left(\frac{r^2W'}{\sqrt{MW}}\right)' - 2wm'\right\}\sin\vartheta,$$

[70] | so that the variation of the quadruple integral

$$\iiiint K\sqrt{g}\; dr\; d\vartheta\; d\varphi\; dl$$

is equivalent to the variation of the single integral

$$\int wm'dr$$

and leads to the Lagrange equations

$$\begin{aligned} m' &= 0 \\ w' &= 0. \end{aligned} \tag{44}$$

It is easy to convince oneself that these equations indeed imply that all $K_{\mu\nu}$ vanish; they therefore represent essentially the most general solution of equations (36) under the assumptions 1., 2., 3., we made. If we take as integrals of (44) $m = \alpha$, where α is a constant, and $w = 1$, which evidently is no essential restriction, then for $l = it$ (43) results in the desired metric in the form first found by Schwarzschild

$$G(dr, d\vartheta, d\varphi, dl) = \frac{r}{r-\alpha}dr^2 + r^2 d\vartheta^2 + r^2 \sin^2\vartheta d\varphi^2 - \frac{r-\alpha}{r}dl^2. \tag{45}$$

The *singularity* of the metric at $r = 0$ disappears only if we take $\alpha = 0$, i.e. *the metric of the pseudo-Euclidean geometry is the only regular metric that corresponds to a world without electricity under the assumptions* 1., 2., 3.

If $\alpha \neq 0$, then $r = 0$ and, for positive α also $r = \alpha$, prove to be places where the metric is not regular. Here I call a metric or gravitational field $g_{\mu\nu}$ *regular* at some place if it is possible to introduce by transformation with unique inverse a coordinate system for which the corresponding functions $g'_{\mu\nu}$ at that place are regular, that is they are continuous and arbitrarily differentiable at the place and its neighborhood, and have a determinant g' that differs from zero.

Although in my view only regular solutions of the basic physical equations represent reality directly, still it is precisely the solutions with places of non-regularity that are an important mathematical instrument for approximating characteristic regular solutions—and in this sense, following Einstein and Schwarzschild, the metric (45), not regular at $r = 0$ and $r = \alpha$, is to be viewed as the expression for | gravity of a [71]
centrally symmetric mass distribution in the neighborhood of the origin[7]. In the same sense a point mass is to be understood as the limit of a certain distribution of electricity about one point, but I refrain at this place from deriving its equations of motion from my basic physical equations. A similar situation prevails for the question about the differential equations for the propagation of light.

Following Einstein, let the *following two axioms* serve as a substitute for a derivation from the basic equations:

The motion of a point mass in a gravitational field is described by a geodesic line, which is a time line[8].

The motion of light in a gravitational field is described by a geodesic null line.

Because the world line representing the motion of a point mass shall be a time line, it is easily seen to be always possible to bring the point mass to rest by *true* spacetime transformations, i.e. there are *true* spacetime coordinate systems with respect to which the point mass remains at rest.

The differential equations of geodesic lines for the centrally symmetric gravitational field (45) arise from the variational problem

---

7 To transform the locations $r = \alpha$ to the origin, as Schwarzschild does, is not to be recommended in my opinion; Schwarzschild's transformation is moreover not the simplest that achieves this goal.

8 This last restrictive addition is to be found neither in Einstein nor in Schwarzschild.

$$\delta\int\left(\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2+r^2\left(\frac{d\vartheta}{dp}\right)^2+r^2\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2-\frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2\right)dp\;=\;0,$$

and become, by well-known methods:

$$\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2+r^2\left(\frac{d\vartheta}{dp}\right)^2+r^2\sin^2\vartheta\left(\frac{d\varphi}{dp}\right)^2-\frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2\;=\;A, \tag{46}$$

$$\frac{d}{dp}\left(r^2\frac{d\vartheta}{dp}\right)-r^2\sin\vartheta\cos\vartheta\left(\frac{d\varphi}{dp}\right)^2\;=\;0, \tag{47}$$

$$r^2\sin^2\vartheta\frac{d\varphi}{dp}\;=\;B, \tag{48}$$

$$\frac{r-\alpha}{r}\frac{dt}{dp}\;=\;C, \tag{49}$$

where A, B, C denote constants of integration. |

[72] I first prove *that the orbits in the $r\vartheta\varphi$- space always lie in planes passing through the center of the gravitation*.

To this end we eliminate the parameter p from the differential equations (47) and (48) to obtain a differential equation for ϑ as a function of φ. We have the identity

$$\begin{aligned}\frac{d}{dp}\left(r^2\frac{d\vartheta}{dp}\right)\;&=\;\frac{d}{dp}\left(r^2\frac{d\vartheta}{d\varphi}\cdot\frac{d\varphi}{dp}\right)\\&=\;\left(2r\frac{dr}{d\varphi}\frac{d\vartheta}{d\varphi}+r^2\frac{d^2\vartheta}{d\varphi^2}\right)\left(\frac{d\varphi}{dp}\right)^2+r^2\frac{d\vartheta}{d\varphi}\frac{d^2\varphi}{dp^2}.\end{aligned} \tag{50}$$

On the other hand, differentiation of (48) with respect to p gives:

$$\left(2r\frac{dr}{d\varphi}\sin^2\vartheta+2r^2\sin\vartheta\cos\vartheta\frac{d\vartheta}{d\varphi}\right)\left(\frac{d\varphi}{dp}\right)^2+r^2\sin^2\vartheta\frac{d^2\varphi}{dp^2}\;=\;0,$$

and if we take from this the value of $\frac{d^2\varphi}{dp^2}$ and substitute on the right of (50), it becomes

$$\frac{d}{dp}\left(r^2\frac{d\vartheta}{dp}\right)\;=\;\left(\frac{d^2\vartheta}{d\varphi^2}-2\cot\vartheta\left(\frac{d\vartheta}{d\varphi}\right)^2\right)r^2\left(\frac{d\varphi}{dp}\right)^2.$$

Thus equation (47) takes the form:

$$\frac{d^2\vartheta}{d\varphi^2}-2\cot\vartheta\left(\frac{d\vartheta}{d\varphi}\right)^2\;=\;\sin\vartheta\cos\vartheta,$$

a differential equation whose general integral is

$$\sin\vartheta\cos(\varphi + a) + b\cos\vartheta = 0,$$

where a and b denote constants of integration.

This provides the desired proof, and it is therefore sufficient for further discussion of geodesic lines to consider only the value $\vartheta = 2/\pi$. Then the variational problem simplifies as follows

$$\delta\int\left\{\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2 + r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2\right\}dp = 0,$$

and the three differential equations of first order that arise from it are |

$$\frac{r}{r-\alpha}\left(\frac{dr}{dp}\right)^2 + r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 = A, \tag{51}$$ [73]

$$r^2\frac{d\varphi}{dp} = B, \tag{52}$$

$$\frac{r-\alpha}{r}\frac{dt}{dp} = C. \tag{53}$$

The Lagrange differential equation for r

$$\frac{d}{dp}\left(\frac{2r}{r-\alpha}\frac{dr}{dp}\right) + \frac{\alpha}{(r-\alpha)^2}\left(\frac{dr}{dp}\right)^2 - 2r\left(\frac{d\varphi}{dp}\right)^2 + \frac{\alpha}{r^2}\left(\frac{dt}{dp}\right)^2 = 0 \tag{54}$$

is necessarily related to the above equations, in fact if we denote the left sides of (51), (52), (53), (54) with [1], [2], [3], [4] respectively we have identically

$$\frac{d[1]}{dp} - 2\frac{d\varphi}{dp}\frac{d[2]}{dp} + 2\frac{dt}{dp}\frac{d[3]}{dp} = \frac{dr}{dp}[4]. \tag{55}$$

By choosing $C = 1$, which amounts to multiplying the parameter p by a constant, and then eliminating p and t from (51), (52), (53) we obtain that differential equation for $\rho = 1/r$ as a function of φ found by Einstein and Schwarzschild, namely:

$$\left(\frac{d\rho}{d\varphi}\right)^2 = \frac{1+A}{B^2} - \frac{A\alpha}{B^2}\rho - \rho^2 + \alpha\rho^3. \tag{56}$$

This equation represents the orbit of the point mass in polar coordinates; in first approximation for $\alpha = 0$ with $B = \sqrt{\alpha}b$, $A = -1 + \alpha a$ the Kepler motion follows from it, and the second approximation than leads to the most shining discovery of the present: the calculation of the advance of the perihelion of Mercury.

According to the axiom above the world line for the motion of a point mass shall be a time line; from the definition of the time line it thus follows that always $A < 0$.

We now ask in particular whether a circle, i.e. $r = \text{const.}$ can be the orbit of a motion. The identity (55) shows that in this case—because of $dr/dp = 0$ —equation (54) is by no means a consequence of (51), (52), (53); the latter three equations therefore are insufficient to determine the motion; instead the necessary equations to be satisfied are (52), (53), (54). From (54) it follows that |

[74]
$$-2r\left(\frac{d\varphi}{dp}\right)^2 + \frac{\alpha}{r^2}\left(\frac{dt}{dp}\right)^2 = 0 \tag{57}$$

or that for the speed v on the circular orbit

$$v^2 = \left(r\frac{d\varphi}{dt}\right)^2 = \frac{\alpha}{2r}. \tag{58}$$

On the other hand, since $A < 0$, (51) implies the inequality

$$r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 < 0 \tag{59}$$

or by using (57)

$$r > \frac{3\alpha}{2}. \tag{60}$$

With (58) this implies the inequality for the speed of the mass point moving on a circle[9]

$$v < \frac{1}{\sqrt{3}}. \tag{61}$$

The inequality (60) allows the following interpretation: From (58) the angular speed of the orbiting point mass is

$$\frac{d\varphi}{dt} = \sqrt{\frac{\alpha}{2r^3}}.$$

So if we want to introduce instead of r, φ the polar coordinates of a coordinate system co-rotating about the origin, we only have to replace

$$\varphi \quad \text{by} \quad \varphi + \sqrt{\frac{\alpha}{2r^3}}t.$$

After the corresponding spacetime transformation the metric

9 Schwarzschild's (loc. cit.) claim that the speed of the point mass on a circular orbit approaches the limit $1/\sqrt{2}$ as the orbit radius is decreased corresponds to the inequality $r \geq \alpha$ and should not be regarded as accurate, according to the above.

$$\frac{r}{r-\alpha}dr^2 + r^2 d\varphi^2 - \frac{r-\alpha}{r}dt^2$$

becomes

$$\frac{r}{r-\alpha}dr^2 + r^2 d\varphi^2 + \sqrt{2\alpha r}\, d\varphi\, dt + \left(\frac{\alpha}{2r} - \frac{r-\alpha}{r}\right)dt^2.$$

| Here the inequality $g_{44} < 0$ is satisfied due to (60), and since the other inequali- [75]
ties (31) are satisfied, *the transformation under discussion of the point mass to rest is a true spacetime transformation*.

On the other hand, the upper limit $1/\sqrt{3}$ found in (61) for the speed of a mass point on a circular orbit also has a simple interpretation. According to the axiom for light propagation this propagation is represented by a null geodesic. Accordingly if we put $A = 0$ in (51), instead of the inequality (59) the result for circular light propagation is the equation

$$r^2\left(\frac{d\varphi}{dp}\right)^2 - \frac{r-\alpha}{r}\left(\frac{dt}{dp}\right)^2 = 0;$$

together with (57) this implies for the radius of the light's orbit:

$$r = \frac{3\alpha}{2}$$

and for the speed of the orbiting light the value that occurs as the upper limit in (61):

$$v = \frac{1}{\sqrt{3}}.$$

In general we find for the orbit of light from (56) with $A = 0$ the differential equation

$$\left(\frac{d\rho}{d\varphi}\right)^2 = \frac{1}{B^2} - \rho^2 + \alpha\rho^3; \tag{62}$$

for $B = \frac{3\sqrt{3}}{2}\alpha$ it has the circle $r = \frac{3\alpha}{2}$ as a Poincaré "cycle"—corresponding to the circumstance that thereupon $\rho - \frac{2}{3\alpha}$ is a double factor of the right-hand side. Indeed in this case—and correspondingly for the more general equation (56)—the differential equation (62) possesses infinitely many integral curves, which approach that circle as the limit of spirals, as demanded by Poincaré's general theory of cycles.

If we consider a light ray approaching from infinity and take α small compared to the ray's distance of closest approach from the center of gravitation, then the light ray has approximately the form of a hyperbola with focus at the center.[10] |

[76] A counterpart to the motion on a circle is the motion on a straight line that passes through the center of gravitation. We obtain the differential equation for this motion if we set $\varphi = 0$ in (54) and then eliminate p from (53) and (54); the differential equation so obtained for r as a function of t is

$$\frac{d^2r}{dt^2} - \frac{3\alpha}{2r(r-\alpha)}\left(\frac{dr}{dt}\right)^2 + \frac{r(r-\alpha)}{2r^3} = 0 \tag{63}$$

with the integral following from (51)

$$\left(\frac{dr}{dt}\right)^2 = \left(\frac{r-\alpha}{r}\right)^2 + A\left(\frac{r-\alpha}{r}\right)^3. \tag{64}$$

According to (63) the acceleration is negative or positive, i.e. gravitation acts attractive or repulsive, according as the absolute value of the velocity

$$\left|\frac{dr}{dt}\right| < \frac{1}{\sqrt{3}}\frac{r-\alpha}{r}$$

or

$$> \frac{1}{\sqrt{3}}\frac{r-\alpha}{r}.$$

For light we have because of (64)

$$\left|\frac{dr}{dt}\right| = \frac{r-\alpha}{r};$$

light propagating in a straight line towards the center is always repelled, in agreement with the last inequality; its speed increases from 0 at $r = \alpha$ to 1 at $r = \infty$.

When α as well as dr/dt are small, (63) becomes approximately the Newtonian equation

$$\frac{d^2r}{dt^2} = -\frac{\alpha}{2}\frac{1}{r^2}.$$

10 A detailed discussion of the differential equations (56) and (62) will be the task of a communication by V. Fréedericksz to appear in these pages.

FROM PERIPHERAL MATHEMATICS TO A NEW THEORY OF GRAVITATION

JOHN STACHEL

THE STORY OF NEWSTEIN OR: IS GRAVITY JUST ANOTHER PRETTY FORCE?

1. 1. INTRODUCTION

In this paper I will argue for the following three theses:

1. The concepts of parallel displacement in Riemannian geometry and of a non-metrical affine connection were developed postmaturely (see Section 2): By the latter third of the nineteenth century, all of the mathematical prerequisites for their introduction were available, and it is a historical accident that they were not developed before the second decade of the twentieth century (see Section 3).
2. The appropriate mathematical context for implementing the equivalence principle is the theory of affine connections on the category of frame bundles, with the bundle morphisms induced by diffeomorphisms on the base manifold (see the Appendix).[1] This theory allows a mathematically precise formulation of Einstein's insight that gravitation and inertia are "essentially the same [*wesensgleich*]" as he put it (see Section 5). The absence of this context constituted a serious obstacle to the development of the general theory of relativity—indeed an insurmountable one to its development by the mathematically most direct route. Consequently, Einstein was forced to take a detour through a long and indirect route from the initial formulation of the equivalence principle in 1907 to the final formulation of the field equations in 1915 (see Section 10). The detour involved focusing attention almost exclusively on the chrono-geometrical structure of spacetime, and to this day, many discussions of the interpretation of the general theory, and of the problem of quantum gravity, still reflect the negative consequences of this detour.
3. Had the concept of an affine connection been developed in a timely manner, the affine formulation of Newtonian gravitation theory, which was actually developed only *after* the formulation of *general* relativity,[2] could have been developed *before* the formulation of *special* relativity. From the outset, such a formulation would have placed appropriate emphasis on the inertio-gravitational structure of

1 Insofar as needed for this paper, these concepts are briefly explained in the Appendix. A particularly useful reference for a more extended discussion of most of these concepts is (Crampin and Pirani 1986).

2 See (Cartan 1923; Friedrichs 1927). Excerpts from Cartan can be found in this volume.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

spacetime and posed the question of its relation to the chronometry and geometry of spacetime (see Sections 6 and 7). When special relativity, with its new chrono-geometry, was developed, this context for gravitation theory would have made the transition from the special to the general theory of relativity rather transparent, thereby avoiding the negative consequences of the actual transition mentioned above.

In order to vivify these rather abstract theses, I have created Isaac Albert Newstein (= Newton + Einstein), a mythical physicist who combines Newton's approach to the kinematical structure of space and time (chronometry and geometry) with Einstein's insight into the implications of the equivalence principle for (Newtonian) gravitation theory (see Section 7). He did this shortly after Hermann Weylmann (= Weyl + Grassmann), an equally mythical mathematician, formulated the concept of affine connection around 1880. Of course, Newstein had to adopt a four-dimensional treatment of space and time in order to carry out his reformulation of Newtonian gravitation theory; but, long before that, the concept of time as a fourth dimension had been introduced in analytical mechanics by d'Alembert and Lagrange.[3]

Continuing my mythical account, when in 1907 Einstein turned to the problem of extending his original (later called special) theory of relativity to include gravitation, Newstein had already shown how to describe the inertio-gravitational field by a non-flat affine connection. Einstein's problem was to combine this insight about the nature of gravitation with the new chrono-geometrical structure of spacetime that he had introduced in 1905. Once the problem is posed in this way, the step from Newstein's formulation of the gravitational field equations to the corresponding equations of Einstein's general relativity is a short one (see Section 8).

Of course, all of this is pure fable; but I believe that—in addition to their entertainment value—such scientific fables are of real value for the history and philosophy of science. First of all, they help us to combat the impression of inevitability often attached to the actual course of historical development, the idea that the "discovery" of a theory is just that: the bringing to light by the intellect of some pre-existing structure, previously hidden but predestined to emerge sooner or later and enter into the scientific corpus in just the form in which it actually did. Secondly, they help us to question the thesis that the formulation of a theory is more-or-less independent of its mode of discovery, including the peculiarities of the individual(s) who happened to "discover" it and the process of negotiation that led to its assimilation into the body of accepted knowledge by the scientific community. Such questions can lead to a more critical re-examination of the current formulation(s) of the theory. We are bound to look more critically at what actually happened, and at the accepted formulation(s) of a theory, if we can produce one or more credible scenarios showing how things might have happened quite differently.[4]

---

3 This is no myth. See my article on "Space-Time," in (Stachel Forthcoming).

2. POSTMATURE CONCEPTS AND THE ROLE OF ABSENCE IN HISTORY

Zuckerman and Lederberg have suggested that, just as there are premature discoveries, "there are postmature discoveries, those which are judged retrospectively to have been 'delayed'" (Zuckerman and Lederberg 1986, 629).[5] I wish to apply the concept of postmaturity to theoretical entities; but since, as noted above, the word "discovery" might suggest a Platonist attitude to mathematical and physical concepts, I shall use more epistemologically neutral phrases: "postmature development," "postmature concept," "postmature theory," etc.

As the work of Zuckerman and Lederberg suggests, in retrospect one can see that—like other forms of absence—the absence of a postmature concept can play a crucial role in the dialectical interplay that shapes the actual course of historical development. My use of word "dialectical" here is purposeful. The second chapter of Roy Bhaskar's book on dialectics (Bhaskar 1993)[6] is entitled: "Dialectic: The Logic of Absences." He equates *absence* with what he calls *real negation*, whose "primary meaning is real determinate absence or non-being (i.e., including non-existence" (Bhaskar 1993, 5). He describes real negation as:

> the central category of dialectic, whether conceived as argument, change or the augmentation of (or aspiration to) freedom, which depends upon the identification and elimination of mistakes, states of affairs and constraints, or more generally ills—argued to be absences alike (Bhaskar 1993, 393).

Elsewhere I shall argue for this viewpoint with examples drawn from the history of music as well as the history of science. But to return to the central concern of this paper, my claim is that "affine connection" is a postmature concept, the absence of which during the course of development of the general theory of relativity had a crucial negative influence on its development and subsequent interpretation. Conversely, the filling of that absence opened the way to a deeper understanding of the nature of gravitation and of its relation to other gauge field theories of physics.

3. A LITTLE HISTORY

Gauss first developed the theory of curved surfaces embedded in Euclidean three-space, including the concepts of intrinsic (or Gaussian) and extrinsic curvature. But he defined these concepts in a way that did not depend on the concept of parallelism.[7] The development of differential geometry had proceeded quite far by the time Rie-

---

4 See (Stachel 1994a) and, for other examples from the history of relativity, (Stachel 1995). For some further comments on alternative histories, see the final section, "Acknowledgements and a Critical Comment."

5 I am indebted to Gerald Holton for drawing my attention to this paper, which fills a gap in my earlier presentations of Newstein's story.

6 I regard Bhaskar's work on critical realism as the most significant attempt at a modern Marxist approach to the philosophy of science (see Stachel 2003a). For a critical introduction to Bhaskar's work, see (Collier 1994).

mann introduced the concept of a locally Euclidean manifold with curvature varying from point to point in 1854, first published in (Riemann 1868).[8] So the idea of starting with a geometrical structure defined in the infinitesimal neighborhood of a point of a manifold and proceeding from the local to the global structure was quite familiar by the last third of the nineteenth century.

Similarly, discussions of the concept of parallelism had played a central role in the development of non-Euclidean geometry in the first half of the nineteenth century.[9] Grassmann's work on affine geometry had abstracted the concepts of parallel lines, plane elements, etc., from their original three-dimensional, Euclidean contexts.[10] Few were aware of the first (1844) edition of the *Ausdehnungslehre*, or even of the second version in 1862; but after the publication of the second edition of the 1844 version in 1878, knowledge of his work began to spread among mathematicians, so that it was widely available to them by the last two decades of the century.[11] By this time, there was already a rich literature on the geometrical interpretation of the principles of mechanics for systems with n- degrees of freedom based on *n- dimensional Riemannian geometry.*[12]

In all this time no one applied Riemann's approach to intervals to the concept of parallelism. Karin Reich has drawn attention to the problem of the delay in the extension of the local approach in geometry to the concept of parallelism:

> Parallelism was and is thus a central theme for the foundations of geometry. Yet it is missing in Bernhard Riemann's Habilitation Lecture "On the Foundational Hypotheses of Geometry," indeed the word parallel does not occur here. Also in the succeeding period of rapidly occurring development of Riemannian geometry parallelism was not a theme. Perhaps this is one of the reasons why Riemannian geometry was not unconditionally accepted by pure geometers (Reich 1992, 78–79).[13]

---

7 Essentially, he defined the intrinsic curvature at a point of a surface in a way that seemed to depend on the embedding of the surface—in terms of the radius of curvature of the sphere that best fits the surface at the point in question—and then proved that the result really does not depend on the embedding. See (Gauss 1902), and for a modern discussion (Coolidge 1940, Book III, chap. III, 355–387).

8 For the history of differential geometry, see (Struik 1933; Coolidge 1940; Laptev and Rozenfel'd 1996, sec. 1: "Analytic and Differential Geometry," 3–26).

9 For the standard older historical-critical account of non-Euclidean geometry, see (Bonola 1955).

10 See (Grassman 1844; 1862; 1878), and for an English translation, (Grassmann 1995). For a survey of publications using Grassmann's approach, demonstrating that their number increased considerably after 1880, see (Crowe 1994, chap. 4); by the end of the century, interest in Grassmann's work was comparable to that in Hamilton's. Weyl was well aware of Grassmann's work. Speaking of affine geometry, he says: "For the systematic treatment of affine geometry with abstraction from the special 3-dimensional case, Grassmann's "Lineale Ausdehnungslehre" (Grassmann 1844)... is the groundbreaking work" (Weyl 1923, 325). In a recent discussion of Grassmann's role as a forerunner of category theory, Lawvere (Lawvere 1996) speaks of "the category A of affine-linear spaces and maps" as "a monument to Grassmann" (p. 255).

11 For a study of Grassmann and his influence, see (Schubring 1996).

12 See (Lützen 1995a; 1995b) for surveys of some of this work.

13 Readers of this work will realize the extent of my indebtedness to Karen Reich's work. I also gratefully acknowledge several helpful discussions with Dr. Reich.

Her retrospective critical judgement a century later is borne out by the contemporary evaluations of those who filled that gap in 1916–1917: Hessenberg, Levi-Civita, Weyl and Schouten.

Hessenberg's paper (Hessenberg 1917) was actually the first such paper, dated June 1916. It starts with a reference to relativity: "Because of the significance that the theory of quadratic differential forms has recently attained for the theory of relativity, the question of whether and how the elaborate and difficult formal apparatus of this theory can be simplified, if not bypassed, gains new significance (p. 187)." Speaking of "Christoffel's well-known transformational calculus," Hessenberg states that his aim is to "replace [it] with a geometrical argument (p. 187)." He criticizes the "formal methods of formation" of various quantities that occur because they do not bring out "the essentially *intuitive* [*anschaulich*] *meaning* of the invariants and covariants needed for the geometrical and physical applications" (p. 191). He stresses the role of Grassmann. "Access [to their geometrical significance] is opened in a way that, to me, seems surprisingly simple by means of Grassmann's ideas" (p. 192).

Levi Civita's paper (Levi Civita 1916), which is dated November 1916, also starts with a reference to Einstein's work:

> Einstein's theory of gravitation ... regards the geometrical structure of space ... as depending on the physical phenomena that take place in it ... The mathematical development of Einstein's magnificent conception ... involves as an essential element the curvature of a certain four-dimensional manifold and the related Riemann symbols [i.e., the curvature tensor] ... Working with these symbols in questions of such great general interest has led me to investigate if it is not possible to simplify somewhat the formal apparatus that is usually used to introduce them and to establish their covariant behavior. Such an improvement is indeed possible ... [This work] started with that sole objective, which little by little grew in order to make room for the geometrical interpretation [of the Riemannian curvature]. At the beginning I thought to have found it in the original work of Riemann ... ; but it is there only in embryo. ... [O]ne gets the impression that Riemann really had in mind that intrinsic and invariant characterization of the curvature, which will be made precise here. On the other hand, however, there is not a trace, either in Riemann or in Weber's commentary, of those specifications (the concept of parallel directions in an arbitrary manifold and consideration of an infinitesimal geodesic quadrilateral with two parallel sides) that we recognize to be indispensable from the geometrical point of view (pp. 173–174).

Reich comments:

> With this word "indispensible" Levi-Cività recalled Luigi Bianchi's characterization of Ricci's absolute differential calculus. Bianchi had characterized this in 1901 as "useful but not indispensable" (Reich 1992, 79–80).

Weyl (1918b) states:

> The later work of Levi-Cività [1916], Hessenberg [1917], and the author [Weyl 1918a][14] shows quite plainly that the fundamental conception on which the development of Riemann's geometry must be based if it is to be in agreement with nature, is that of the infinitesimal parallel displacement of a vector.[15]

14 For a discussion of this and the succeeding editions of Weyl's book, see the next section.

After the introduction of Riemannian parallelism by Hessenberg and Levi-Civita (and, again independently in (Schouten 1918)), it was but a brief and natural step to its generalization. Since the abstraction (in the large) of affine parallelism from parallelism in Euclidean geometry had already been made, the abstraction (in the small) of affine parallelism from parallelism in a Riemannian manifold is immediately suggested by the analogy. Indeed, Weyl took that step just a year later: In (Weyl 1918a) he defines an affinely connected manifold.[16]

The evidence thus indicates that both the Riemannian concept of parallelism and its affine generalization were introduced *postmaturely*. The absent concept of Riemannian parallelism could have been filled at any time during the last third of the nineteenth century, and followed quickly by the introduction of the concept of an affinely connected manifold, since it is a natural generalization of the Grassmannian "*lineale Ausdehnungslehre*."

Indeed, Grassmann himself might have accomplished these tasks. Towards the end of his life he learned about the work of Riemann and Helmholtz, and one of his last publications (Grassmann 1877) discusses the relation of their work to his *Ausdehnungslehre*. He discusses a method of introducing such non-linear geometries that amounts essentially to defining them as subspaces of linear spaces of higher dimensions. The path that Levi-Civita initially took to the definition of Riemannian parallelism was based on embedding a Riemannian space in a Euclidean space of sufficiently high dimension. Had Grassmann lived longer, it is conceivable that he might have introduced the concept of affine parallelism by a similar method (see the discussion in the Appendix). But he died in the same year that he wrote this paper; so I have been forced to invent Weylmann, the mathematician who introduces the concept of an affinely connected manifold around 1880, neither prematurely nor postmaturely.

4. EQUIVALENCE PRINCIPLE AND AFFINE CONNECTION

It was Albert Einstein who first realized the profound significance of the equality of inertial and gravitational mass. He soon began to speak of inertia and gravitation as "*wesensgleich*": essentially the same in nature. By an acceleration of the frame of reference, the division between inertial and gravitational "forces" can be altered, and indeed by a suitably chosen acceleration the combination of both can even be made to vanish at any point of spacetime.

Einstein's problem was to find the way to incorporate this physical insight into the mathematical structure of gravitation theory. After the development of the concept of affine connection, the way became clear: there is an inertio-gravitational field, repre-

15 Translated from (Weyl 1923, 202).

16 For references and discussion of the work of Levi-Civita, Hessenberg, Schouten and Weyl, see the indispensible (Reich 1992). For the background to Weyl's "Purely Infinitesimal Geometry," see (Scholz 1995). I am indebted to Dr. Erhard Scholz for a discussion of this work.

sented mathematically by a symmetric connection in spacetime, which incorporates this essential unity in its very nature. We can see the development of this insight by looking at the various editions of Weyl's *Raum-Zeit-Materie*. In (Weyl 1918b), Levi-Civita's concept of parallel transport, based upon the embedding of a Riemann space in a flat Euclidean space of higher dimension, is freed from this dependence by giving it an intrinsic definition. Weyl further states that the Christoffel symbols represent the gravitational field. In (Weyl 1919)—which follows the argument of Weyl (1918a)—the concept of parallel transport is freed from its dependence on the metric field by the introduction of the concept of affine connection. Weyl (1921) refers to this connection as the "guiding field" (*Führungsfeld*), incorporating the effects of both gravitation and inertia on the motion of bodies.

Soon afterwards, Cartan (1923) drew the obvious conclusion: By incorporating the equivalence of gravitation and inertia into Newton's gravitation theory, it can be formulated in terms of a Newtonian affine connection. Since then, starting with (Friedrichs 1927) and culminating—but certainly not ending—in (Ehlers 1981), a series of refinements of Cartan's approach have brought the affine version of Newton's theory to a state of considerable mathematical perfection.

However, I shall not give the most, abstract, coordinate-free characterization of the Newtonian affine connection based on the simplest set of axioms. For our purposes, it will be more useful to show how, starting from the usual form of the Newtonian theory of gravitation, the components of the connection with respect to a physically chosen basis may be defined, thus suggesting how Newstein could have proceeded—had he only existed![17]

5. NEWSTEIN'S WORLD

We shall start from the usual formulation of Newtonian gravitation theory in some inertial frame of reference (ifr, for short). Events in this frame are individuated with the help of the Newtonian absolute time t (chronometry), and three Cartesian coordinates (i.e, assuming Euclidean geometry), fixed relative to some choice of origin O and of three mutually perpendicular axes.[18] Since inertial and gravitational mass are equal, if $\boldsymbol{g}$ represents the force/unit gravitational mass, the equation of motion of a (structureless) particle will be

$$\boldsymbol{a} = \boldsymbol{g}, \tag{1}$$

where $\boldsymbol{a}$ is the acceleration of the particle with respect to the chosen ifr.

---

17 See (Stachel 1994b) for a somewhat more abstract discussion of spacetime structures in Newton-Galilean and special-relativistic spacetimes (i.e., in the absence of gravity), and in Newtonian and Einsteinian gravitational theories.

18 We assume units of time and distance fixed initially and used in all frames of reference, and shall use vector notation, so that, for example, the displacement vector from the origin $\boldsymbol{r} = (x^1, x^2, x^3)$, the velocity $\boldsymbol{v} = d\boldsymbol{r}/dt$, the acceleration $\boldsymbol{a} = d\boldsymbol{v}/dt$, etc, all with respect to the ifr, are denoted by boldface symbols.

Now consider transformations to another frame of reference, moving linearly with respect to the first:

$$\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{R}(t), \quad t' = t. \tag{2}$$

If the velocity vector $\boldsymbol{V} = d\boldsymbol{R}/dt$ is *constant*, then the transformation is to another inertial frame of reference, and the equation of motion, eq. (1), is invariant under such a transformation. That is, both $\boldsymbol{a}$ and $\boldsymbol{g}$ are invariant under such *Galilei transformations* from one inertial frame to another.

However, if $\boldsymbol{V}$ is *not* constant, then the transformation is to some linearly accelerated (rigid) frame of reference, and differentiation of eq. (2) twice with respect to the time gives

$$\boldsymbol{a}' = \boldsymbol{a} - \boldsymbol{A}(t), \quad \boldsymbol{A}(t) = d^2\boldsymbol{R}(t)/dt^2. \tag{3}$$

In Newtonian mechanics, "true" forces, such as $\boldsymbol{g}$, are assumed to be the same in all frames of reference. To compensate for the use of a non-inertial frame of reference, so-called "inertial forces" appear in the equations of motion (such forces might better be called "non-inertial"). Indeed, when we substitute eq. (3) in eq. (1), we get:

$$\boldsymbol{a}' + \boldsymbol{A} = \boldsymbol{g}, \text{ or } \boldsymbol{a}' = \boldsymbol{g} - \boldsymbol{A}, \tag{4}$$

and $-\boldsymbol{A}(t)$ appears as such an "inertial force" in the equation of motion of a particle with respect to a linearly accelerating frame.

But, one may ask, if we carry out measurements in some frame of reference, and get an acceleration, let us say $\boldsymbol{a}'$, for a test particle, how do we separate it into its components, the "true force" $\boldsymbol{g}$ and the "inertial force" $-\boldsymbol{A}$? Newton would not have hesitated a moment in answering: Look for the sources of the gravitational force, and use the inverse square law to compute the total $\boldsymbol{g}$ at the point where the test particle is located. Alternatively, he might have proposed: Look at the center of mass of the "system of the world" (i.e., the solar system) and see whether you are accelerating relative to it to find $\boldsymbol{A}$.

But by the end of the nineteenth century, under the influence of Maxwell's electromagnetic theory, the field point of view towards forces was beginning to prevail; according to this viewpoint, one should look upon the gravitational field as the conveyor of all gravitational interactions between massive bodies. Accordingly, the local gravitational field at a point in space (and an instant of time) should always be ascertainable by means of local measurements in the neighborhood of that point. Now, in the case of any other force but the gravitational, there would be no obstacle to separating out the inertial from the non-gravitational effects. For electrically charged particles, for example, one would merely vary the ratio of electric charge to inertial mass: The electric force would vary with this ratio, the inertial force would not. But the ratio of gravitational charge (= gravitational mass) to inertial mass is just what *cannot* be varied—the invariance of that ratio is the primary empirical basis of the equivalence principle.

So the answer to our question is: Once we adopt the field point of view about gravitation, there is no way (locally) to distinguish inertial from gravitational effects.

We have to recognize that there is an *inertio-gravitational field*, and that how this field divides up into inertial and gravitational terms is not absolute (i.e., frame-independent), but depends on the state of motion (in particular the acceleration) of the frame of reference being used. Indeed, we see that, by choosing the value of $\boldsymbol{A}$ to coincide numerically with the value of $\boldsymbol{g}$ at some point, we can make the total inertio-gravitational field vanish at that point. Indeed, this is why we did not call it an inertio-gravitational force: Although the values of their components with respect to some frame of reference can change depending on the state of motion of that frame, non-vanishing force fields at a point, such as the electric and magnetic fields making up the electromagnetic field, cannot be made to all vanish by any change of reference frame.

Another consequence of our new, equivalence-principle viewpoint is that a basic distinction between inertial and linearly accelerated frames of reference is no longer tenable. Any rigid non-rotating frame of reference is just as good as any other.

Let us now inventory what is left after we adopt this new viewpoint:

1. the absolute time, assumed to be measurable by ideal clocks; its measurement is unaffected by the presence of an inertio-gravitational field (compatibility of chronometry with the inertio-gravitational field);
2. Euclidean geometry, which holds within each frame in the class of three-dimensional, non-rotating frames of reference; it is assumed to be measurable with ideal measuring rods; its measurement is unaffected by the presence of the inertio-gravitational field (compatibility of geometry with the inertio-gravitational field).
3. Since gravitation and inertia are no longer (absolutely) distinguished (i.e., gravity is no longer regarded as a force), the set of "force-free" inertial motions is replaced by a set of "force-free" inertio-gravitational motions. One of these is determined by specifying a velocity vector at a point of space and an instant of time. The vector is then the tangent to the "freely falling motion" through the point at this instant.
4. While the inertio-gravitational field $\boldsymbol{g}(\boldsymbol{r}, t)$ is not absolute (i.e., it depends on the frame of reference used, and only behaves like a vector with respect to transformations within a given frame of reference), its spatial derivatives $\partial_m g^n(\boldsymbol{r}, t)$ are independent of the (non-rotating) reference frame. Physically, these differential gravitational forces are usually designated as the tidal forces, since they are responsible for the tides, among other effects. The matrix of these quantities determines the relative acceleration of two freely falling test particles, i.e., the acceleration of one particle with respect to the other. The components of the tidal forces therefore may be evaluated by measurement of the components of this relative acceleration.

6. THE NEWTONIAN CONNECTION

Now we are ready to make the transition to the four-dimensional point of view, in which a point of spacetime is specified by the four coordinates (t, x^1, x^2, x^3) or $(t, \boldsymbol{r})$ for short, where x^1, x^2, x^3 are the Cartesian coordinates of the point with respect to some non-rotating frame of reference and t is the absolute time.[19] We shall refer to these as *adapted coordinates* for this frame of reference. The absolute time gives a *foliation* of spacetime, i.e., a family of non-intersecting hypersurface that fills the spacetime. In the adapted coordinate system the foliation consists of the hypersurfaces $t = \text{const}$. A vector is said to be *space-like* if it is tangent to a hypersurface of the foliation; a vector is *time-like* if it is not space-like. Any curve, the tangent vector to which is always time-like, is a *time-like* curve, with a similar definition for space-like curves. In adapted coordinates a vector is time-like if it has a non-vanishing time component, space-like if it does not.

We can use any (three-)velocity field $\boldsymbol{v}(t)$ to rig the hypersurfaces of constant time: Define a time-like four-velocity field $V(t)$, with t-component = 1 and spatial components equal to those of $\boldsymbol{v}(t)$ in adapted coordinates. Thus, $V(t)$ defines a congruence of time-like curves that fills spacetime. Indeed, we need merely select *one* such time-like curve $V(t)$ and then parallel propagate it along each hypersurface $t = \text{const}$ to get this congruence. In particular, the paths of the points $x^1, x^2, x^3 = \text{const}$, parametrized by the absolute time t, constitute such a congruence; Euclidean geometry holds for these spatial coordinates at all times. Thus we have specified the chronometry and the geometry of the initial frame of reference using the adapted coordinates.

Any $V(t)$ field provides a rigging of each hypersurface (see the discussion of rigged hypersurfaces in the Appendix). Just as a rigging was needed to go from the flat affine connection of the enveloping space to the non-flat affine connection of a hypersurface embedded in it, a rigging is needed here to relate the flat (Levi-Civita) connection on each Euclidean hypersurface to the four-dimensional non-flat connection that we want to define for spacetime as the mathematical representation of the inertio-gravitational field.

Indeed, we can define a unique symmetric, four-dimensional affine connection on the spacetime by requiring that it satisfy the following conditions:

1. The absolute time is the affine parameter for all time-like geodesic paths. A geodesic path that is time-like at any of its points is time-like at all its points.
2. There is a flat, Euclidean connection on each (three-dimensional) hypersurface of the foliation. Hence, the Euclidean distance is the affine parameter for each space-like geodesic path. A geodesic path that is space-like at any of its points is space-like at all its points.

---

19 We shall designate a time component by a sub- or superscript "t," and spatial components by sub- or superscript "$i, j, k\ldots$" or other lower-case Latin letters.

3. The three-dimensional and the four-dimensional treatments of the spatial geometry on each hypersurface are consonant with each other: The Euclidean (flat) three-dimensional affine connection on each hypersurface of some frame of reference coincides with the connection induced on that hypersurface by the four-dimensional connection when that hypersurface is rigged with *any* time-like $V(t)$ field.[20]
4. Parallel transport of any space-like vector is path-independent. By picking an orthonormal triad $\boldsymbol{e}_i$ of such vectors at some point on an initial hypersurface of the foliation, and parallel transporting the triad along any time-like curve with tangent vector $V(t)$, a frame of reference is generated: Once it is parallel transported to a point on another hypersurface of the foliation, the triad can be propagated to any other point of the hypersurface by (path-independent) parallel transport.
5. If we add any $V(t)$ to the triad field e_i, now interpreted as four-vectors, we get a four- dimensional frame of reference.[21] In any such frame of reference, any path that obeys the Newtonian gravitational equation of motion of a structureless test particle shall be a time-like geodesic of the four-dimensional connection parametrized by the absolute time. The spatial projection of its four-dimensional tangent vector onto any hypersurface of the foliation will coincide with the three-velocity of the test particle on that hypersurface.

As indicated earlier, we have not attempted to give a minimal list of assumptions, each of which is independent of the others; but rather, a physically intuitively plausible list. We now proceed to derive the components of the connection in some given non-rotating frame of reference, i.e., using coordinates adapted to the tetrad of basis vectors $V(t)$, e_i that characterize this frame of reference.

The equation of a geodesic in these coordinates is (see the Appendix):

$$d^2x^\kappa/d\lambda^2 + \Gamma^\kappa_{\rho\sigma}(dx^\rho/d\lambda)(dx^\sigma/d\lambda) = 0, \quad (\kappa, \rho, \sigma = t, x^1, x^2, x^3), \tag{5}$$

where λ is an affine parameter, i.e., the (four-dimensional) tangent vector to the curve $P(\lambda)$ is equal to $dP/d\lambda$; and the components of the connection are with respect to the chosen four-dimensional frame of reference. If we consider time-like geodesics, condition 1) requires that t be an affine parameter for all of them. The four-velocity $dx^\rho/d\lambda$ will thus have components $(1, \boldsymbol{v})$ in the adapted coordinate system, where $\boldsymbol{v}$ is the three-velocity of the particle. Considering only the t- component of eq. (5) for the moment, in adapted coordinates it takes the form:

---

20 If the requirement is fulfilled for one such field it is fulfilled for any such field, since two such fields can only differ by a *space-like* acceleration vector field. So the transition from one non-rotating frame of reference to another, which corresponds mathematically to a change of $V(t)$ field, does not affect the result.

21 Note that any such $V(t)$ field commutes with the three e_i fields, which commute with each other, so that they form a holonomic basis; so coordinates adapted to this basis will always exist.

$$d^2t/dt^2 + \Gamma^t_{tt} + 2(\Gamma^t_{ti})v^i + (\Gamma^t_{ij})v^iv^j = 0, \tag{5a}$$

and since the first term vanishes, the only way that eq. (5a) can hold for all values of v^i is if Γ^t_{tt}, Γ^t_{ti} and Γ^t_{ij} all vanish in the adapted coordinate system. In other words, these are the mathematical conditions that assure the compatibility of the chronometry and the inertio-gravitational field. Physically, this means that an ideal clock moving around in the inertio-gravitational field will always measure the absolute time.

Conditions 2, 3, and 4 now demand that the three space-like vectors e_i, which lie along the coordinate axes and thus have components δ^μ_i in adapted coordinates, have vanishing covariant derivates with respect to both the Euclidean (flat) three-dimensional connection on each hypersurface, and the non-flat inertio-gravitational four-dimensional connection. By a similar argument to that above, these conditions result in the vanishing of Γ^m_{ti} and Γ^m_{ni} in the adapted coordinate system. In other words, these are mathematical conditions that assure the compatibility of the geometry and the inertio-gravitational field. Physically, this means that an ideal measuring rod moving around in the inertio-gravitational field will always measure the Euclidean distance.

Condition 5 now fixes the values of the only remaining non-vanishing components of the affine connection, Γ^m_{tt}, in the adapted coordinate system. Returning to eq. (5), its spatial components in the adapted coordinate system now take the form:

$$d^2x^m/dt^2 + \Gamma^m_{tt} = 0, \tag{5b}$$

all other terms in the equation vanishing because of the previously-established vanishing of the other components of the connection. We see that we need merely set:

$$\Gamma^m_{tt} = -g^m(\boldsymbol{r}, t) \tag{6}$$

in the adapted coordinates in order to have the geodesic equation coincide with the equation of motion of a particle in the gravitational field $\boldsymbol{g}(\boldsymbol{r}, t)$.

We have now fixed all the components of the symmetric affine connection in the adapted coordinate system. We need merely apply the general transformation law for the components of the connection under a coordinate transformation $x^{\kappa'} = x^{\kappa'}(x^\kappa)$:

$$\Gamma'^\kappa_{\mu\nu} = \Gamma^\kappa_{\mu\nu}(\partial x^{\kappa'}/\partial x^\kappa)(\partial x^\mu/\partial x^{\mu'})(\partial x^\nu/\partial x^{\nu'}) + \partial^2 x^{\kappa'}/\partial x^\mu \partial x^\nu, \tag{7}$$

to the equations for a linearly accelerated transformation (see eq. (2) of Section 5):

$$x^{m'} = x^m + R^m(t), \tag{8}$$

in order to see that the components of the connection transform correctly; i.e, that all the components but Γ'^m_{tt} continue to vanish, and the Γ^m_{tt} transform just like the components of $\boldsymbol{g}$ under such a transformation (see Section 5, eq. (4)). If we carry out a transformation to a rotating system of coordinates, the transformation of the components of the connection introduces terms that correspond to the Coriolis and centripe-

tal "inertial forces" that must be introduced in a rotating coordinate system. To get the form of the components of the connection in an arbitrary coordinate system, one need merely apply eq. (7) to an arbitrary coordinate transformation.

What about the tidal forces, which as mentioned above are absolute? They are represented by the appropriate components of the Riemann tensor, which can be computed from the Newtonian inertio-gravitational connection. Since they are components of a tensor, they indeed possess an absolute character, in the sense that if the components do not all vanish at a point, no change in frame of reference at that point can make them all vanish. These components of the Riemann tensor enter into the equation of geodesic deviation, which describes in four-dimensional tensorial form the relative acceleration of two particles falling freely in the inertio-gravitational field; but I shall not enter into details here.

Rather, I turn to the question of the field equations for the inertio-gravitational field. The Newtonian field $\boldsymbol{g}$ obeys the field equation:

$$\nabla \cdot \boldsymbol{g} = 4\pi G\rho, \tag{9}$$

where G is the Newtonian gravitational constant, ρ is the mass density of the material sources of the gravitational field, and $\nabla \cdot \boldsymbol{g}$ is the trace of the tidal force matrix. If one works out the components of $R_{\mu\nu}$, the contracted Riemann or Ricci tensor, in the adapted coordinates, it turns out that only R_{tt} is non vanishing, and it equals $-\nabla \cdot \boldsymbol{g}$. So $R_{tt} = -4\pi G\rho$ and all other components =0 in the adapted coordinates. The only remaining problem is to write this result as a tensorial equation, independent of coordinate system; but this is easily solved by introducing a covariant vector field T_μ, such that in adapted coordinates $T_\mu = \partial_\mu t = \delta^\mu_t$. The gravitational field equations now take the tensorial form:

$$R_{\mu\nu} = 4\pi G\rho T_\mu T_\nu, \tag{10}$$

which is clearly of the same form in all coordinate systems.

In a more complete treatment,[22] one would have to go a step further: the Newtonian gravitational field $\boldsymbol{g}$ can be derived from a gravitational potential function ϕ: $\boldsymbol{g} = -\nabla\phi$, and this condition can be expressed intrinsically in terms of the properties of the corresponding Riemann tensor (the tidal force matrix introduced in Section 5, which is closely related to certain components of the Riemann tensor, becomes symmetric). Now ϕ plays an important role in taking the Newtonian limit of general relativity, but since we shall not discuss this issue, I can forego entering into further consideration of details.

The non-dynamical Newtonian chrono-geometrical structures, consisting of the absolute time and the relative spaces of the family of non-rotating frames of reference, are unmodified by the presence of gravitation. Mathematically, they are represented by a closed temporal one-form (the T_μ introduced above) and a trivector field

22 See (Stachel 2003b) for such a treatment.

whose transvection with the one-form vanishes (the e_i introduced above), from which a degenerate (rank 3) spatial "metric" may be constructed ($=\delta^{ij}e_i e_j$).

However, the compatible flat inertial structure of Newton's theory is modified. It becomes a dynamical structure, the Newtonian inertio-gravitational field, which remains compatible with the chrono-geometrical structures. Mathematically, it is represented by a symmetric affine connection (the Newtonian connection $\Gamma^{\kappa}_{\rho\sigma}$ discussed above), which can be derived from a "connection potential" (the ϕ discussed above). Its contracted Riemann tensor obeys field equations that relate it to the masses acting as its source (eq. (10) above). The compatibility of this connection with the chrono-geometrical structure means, as noted earlier, that clocks and measuring rods freely falling in the inertio-gravitational field still measure absolute temporal and spatial intervals, respectively. Mathematically, this is expressed by the vanishing of the covariant derivatives of the temporal one-form and degenerate spatial "metric" with respect to the Newtonian connection.

7. SOME MYTHICAL HISTORY: NEWSTEIN MEETS WEYLMANN

Once the concept of affine connection has been developed and the Riemann tensor geometrically interpreted in terms of parallel transport around closed curves, this version of Newton's theory—which converts gravitation from a force that pulls bodies off their (non-dynamical) inertial paths, into a (dynamical) modification of the (inertial) affine connection—is almost immediately suggested by the equality of gravitational and inertial mass. Indeed, shortly after the mythical mathematician Weylmann formulated the concept of affine parallelism, his equally mythical physicist colleague Newstein developed this reinterpretation of Newtonian gravitational theory. Brooding on the equality of gravitational and inertial mass, he became convinced of the essential unity of gravitation and inertia. Originally, he expressed this insight in the usual three-plus-one language of physics, treating space and time separately (see Section 5). He considered uniformly accelerated frames of reference in the absence of gravitation (the Newstein elevator!), and decided it was impossible to distinguish such a frame of reference from a non-accelerated frame with a constant gravitational field. This led him to consider transformations between linearly accelerated frames of reference.

He was puzzled by the strange transformation law that he had to introduce for the gravitational "force," which no longer behaves like a vector under such transformations. At some point he turned to Weylmann, who soon realized that the gravitational "force" transforms like the Γ^{m}_{tt} components of a four-dimensional affine connection, and that Poisson's law for the gravitational potential could be written as an equation linking the Ricci tensor of the connection with its material sources (see Section 6). In the now-famous Newstein-Weylmann paper, the two developed a four-dimensional geometrized formulation of Newtonian gravitation theory, which generalized Newtonian chrono-geometry to include linearly accelerated frames and a dynamized inertio-gravitational connection field, but still included the concept of absolute time.

In so far as they took any notice of this work, their contemporaries regarded it as an ingenious mathematical *tour-de-force*. But, since it had no new physical consequences, it did not much impress Newstein's positivistically-inclined physics colleagues.

Weylmann analyzed the invariance group of the new theory, which is much wider than that of the older Newtonian kinematics. The privileged role of the inertial frames of reference in Newton's theory, just beginning to be realized thanks to the work of Lange and Neumann, was lost in the new interpretation of gravitation.While rotation remained absolute (in the sense that all components of the connection representing centrifugal and Coriolis forces could be made to vanish globally by a coordinate transformation), all linearly accelerated frames of reference were now equal, and the significance of this occasioned a discussion among a few philosophers of science who concerned themselves with the foundations of mechanics. Ernst Mach added a few lines about Newstein to the latest edition of his *Mechanik*.

8. MORE MYTH: EINSTEIN CONFRONTS NEWSTEIN

Perhaps this is where Albert Einstein first read about Newstein's work. At any rate, in 1907, pursuant to his commission to write a review article on the physical consequences of his 1905 work on the relativity principle (now becoming known as the theory of relativity),[23] he turned his attention to gravitation, and (like Newstein) was struck by the equality of gravitational and inertial mass. He realized that, as a consequence, in Newtonian mechanics there is a complete equivalence between an accelerated frame of reference without a gravitational field and a non-accelerated frame of reference, in which there is a constant gravitational field. He soon generalized this to what he later called the principle of equivalence: There is no physical difference (mechanical or otherwise) between the two frames of reference.[24]

Recalling what he had read about Newstein, Einstein realized that he had rediscovered the loss of the privileged role of inertial frames once gravitation is taken into account. Like Newstein, he became convinced that inertia-cum-gravitation must be represented mathematically by an affine connection; but now this representation somehow must be made compatible with the new chronogeometry he had developed in his 1905 theory.[25] He first tried to preserve the non-dynamical nature of this chrono-geometrical structure—which Minkowski soon expressed in terms of a four-dimensional pseudo-Euclidean geometry—by developing various special-relativistic gravitational theories that incorporated the unity of gravitation and inertia by the very fact that they were based upon an affine connection. But the Riemann tensor of the inertio-gravitational connection in each of these theories was non-vanishing, while

23 For a translation of this paper, see (Stachel 1998).

24 Aside from the first sentence, this paragraph is a summary of the actual historical circumstances of Einstein's first work on gravitation, see (Einstein 1907). The fantasy begins in the next paragraph.

25 In the frame bundle language, the physically preferred subgroup of the general linear group had to be changed from the Newtonian group to the Lorentz group.

the metric-affine structure of Minkowski spacetime is flat. Physically, this meant that the inertio-gravitational and chrono-geometrical structures were not compatible: Good clocks and measuring rods, as defined by the chrono-geometrical structure, did not keep the proper time or measure the proper length when moved about in the gravitational field.

While this could be "explained away" as due to a universal distorting effect of gravitation on all measuring rods and clocks, something about such an explanation disturbed him. Since the effect was universal, the "true" Minkowski chronogeometry could be shown to have no physically observable consequences.

Finally, he realized what was bothering him: This type of explanation was all too similar to Lorentz's interpretation of the Lorentz transformations: Galilean chronogeometry is the "true" one; but the universal effect of motion through the absolute (aether) frame of reference exerts a universal effect on all physical processes that prevents any physically observable consequences of this motion. What was the way out of this new unobservability dilemma?

Suddenly the answer struck him: If he required compatibility between the inertio-gravitational and chrono-geometrical structures, the problem would disappear, just as it had in Newstein's reinterpretation of Galilean kinematics. Good measuring rods and clocks, as defined by such a chrono-geometrical structure, would measure the true proper lengths and times wherever they were placed in the inertio-gravitational field. But there was a price to pay for this compatibility: The chrono-geometrical field could no longer be flat. It would have a curvature attached to it in the Gaussian sense, the one that Riemann originally had generalized from two to an arbitrary number of dimensions. In this theory, the Riemann tensor would have two distinct (but compatible) interpretations: as the curvature of a connection, associated with parallel transport and the equation of geodesic deviation; and as the curvature of a pseudo-metric, associated with the Gaussian curvature of each of the two-dimensional sections at any point of spacetime.

And of course, since metric and connection were now compatible, this implied that the components of the connection with respect to any basis were numerically equal to the Christoffel symbols of the metric with respect to that basis. And since the connection is a dynamical field, the metric would also have to become a dynamical field. In contrast to the Newsteinian case, where the chrono-geometry remained non-dynamical, in the Einsteinian case, there are no non-dynamical spacetime structures. The bare manifold remained absolute in a certain sense;[26] but then, it had no physical characteristics other than dimensionality and local topology unless and until the iner-

26 I say this because, in actual fact, the global topology of the manifold is not given before the metric-cum-connection field, as implied in so many presentations of general relativity. One actually solves the Einstein field equations on a small patch, and then looks for the maximal extension of that patch compatible with the given metric. Certain criteria for compatibility must be given before the question of maximal extension(s) becomes meaningful, of course. For discussion of this topic, see (Stachel 1986; 1987).

tio-gravitational cum chronogeometrical field was impressed upon it. Least of all do the points of the manifold represent physical events before imposition of a metric.[27]

The new, dynamical theory of spacetime structures had a number of novel physical consequences, and Einstein soon became world-famous—but you know the rest of the story.

9. SOME REAL HISTORY: EINSTEIN WITHOUT NEWSTEIN

Unfortunately, the last section was a historical fable, and the real Einstein had to work out the general theory of relativity in the absence of the concept of affine connection—an absence which, as suggested in Section 2, played a fateful role in the actual development and subsequent history of the theory. It took Einstein without Newstein seven years to develop the general theory of relativity *after* he had adopted the equivalence principle as the key to a relativistic theory of gravitation. Rather than tell the entire story of the many genial steps and equally numerous missteps on Einstein's road from special to general relativity,[28] I shall here just highlight some of the most fateful consequences of the absence of the connection.

First of all, it is important to realize that the tensor calculus, as originally developed by Christoffel, Ricci, Levi Civita and others, was a branch of invariant theory, with only tenuous ties to geometry.[29] Einstein's introduction of the metric tensor field as the mathematical representation of both the chrono-geometry of spacetime and the potentials for the gravitational field did not carry with it most of the geometrical implications that we take for granted today. Insofar as it did carry geometrical implications, notably in fixing the geodesics of the manifold, this had to do with the interpretation of geodesics as the shortest paths (or rather longest, for time-like paths—the twin paradox) in spacetime. The interpretation of geodesics as the straightest paths in spacetime, more important for the understanding of the gravitational field—in particular, the interpretation of the Riemann tensor in terms of the equation of geodesic deviation—had to await the work of Levi Civita and Weyl on parallelism discussed in Section 3.[30] Curvature, in other words, was given the Gauss-Riemann interpretation, rather than the interpretation as the tendency of geodesics to coverge (or diverge), leading to its association with tidal forces.

---

27 For discussion of the hole argument, which bears on this point, see (Stachel 1993) and references therein.

28 See the first two volumes of this series on the development of general relativity. For earlier accounts by this author and others, see (Stachel 1995) and the references therein.

29 "The calculus developed by Gregorio Ricci in the years 1884–1887 had its roots in the theory of invariants, therefore it naturally lacked a geometrical outlook or interpretation, and was so intended by Ricci" (Reich 1992, 79). For the history of the tensor calculus, see (Reich 1994).

30 Interestingly, this interpretation was anticipated by Hertz in his geometrical version of mechanics. See (Hertz 1894) and, for a discussion of the 19th century tradition of geometrical interpretations of mechanics, (Lützen 1995a; 1995b).

It is often said that Einstein, with the help of Grossmann, found ready-to-hand the mathematical tools he needed to develop general relativity: Riemannian geometry and the tensor calculus. But this statement must be taken with a large grain of salt. It would be more correct to say that he had to make do with the tools at hand, with important negative consequences for the development of the theory, and—more importantly for us now—with negative consequences for the interpretation of the theory that continue to exert their effects to this day.[31]

To give two concrete examples of this negative influence on Einstein's work:

1. Until late in 1915, he regarded the derivatives of the metric tensor, rather than the Christoffel symbols, as the mathematical representative of the gravitational-cum-inertial field.[32] In Einstein 1915, he finally corrected this error:

   > These conservation laws [the vanishing of the covariant derivative of the stress-energy tensor] previously misled me into regarding the quantities $1/2\sum_{\mu} g^{\tau\mu}\partial g_{\mu\nu}/\partial x_{\sigma}$ as the natural expression for the components of the gravitational field, although in the light of the formulas of the absolute differential calculus it seems more obvious to introduce the Christoffel symbols instead of these quantities. This was a fateful prejudice (Einstein 1915, 782).

   The reason why this error was so fateful is that it mislead Einstein in his search for the gravitational field equations, a search that took over two years *after* he had adopted the metric tensor field as the mathematical representation of gravity.[33]

2. From 1912 onwards, Einstein expected that, in the Newtonian limit of general relativity, the spatial part of the metric field tensor would remain flat and that the g_{oo} component of the metric would reduce to the Newtonian gravitational potential. Correctly understood, in terms of a formulation of the theory taking the Newtonian limit of both the connection and the metric, these expectations are fulfilled. But one cannot properly take the Newtonian limit of general relativity without the concept of an affine connection, and the corresponding affine reformulation of Newtonian theory discussed in Section 6. Indeed, the problem of correctly taking the Newtonian limit of general relativity only began to be solved in (Friedrichs 1927), and the process was not completed in all details until (Ehlers 1981). In the absence of the affine approach, more-or-less heuristic detours through the weak-field, fast motion (i.e., special-relativistic) limit followed by a slow motion approximation basically out of step with the fast-motion approach, had to be used to "obtain" the desired Newtonian results.[34]

---

31 Perhaps the first such negative influence on work done after the final formulation of the general theory is the ultimate failure of Lorentz's attempt to give a coordinate-free geometrical interpretation of the theory. I thank Dr. Michel Janssen for pointing this out to me. For an account of Lorentz's attempt, see (Janssen 1992).

32 See (Einstein and Grossmann 1913, 7), and (Einstein 1914, 1058), for examples.

33 For details see vol. 1 of this series on the development of general relativity.

34 See (Stachel 2003b) for more details.

Einstein originally thought that he knew the form of the weak field metric in the static case. It involved a spatially flat metric tensor field, with only the g_{oo} component of the metric depending on the coordinates. He used this form of the static metric as a criterion for choosing the gravitational field equations: This form of the metric had to satisfy the field equations, which led to a disastrous result: No field equation based on the Ricci tensor had this form of the static metric as a solution, and Einstein abandoned the Ricci tensor for over two years![35] Had he known about the connection representation of the inertio-gravitational field, he would have been able to see that the spatial metric can go to a flat Newtonian limit, while the Newtonian connection remains non-flat without violating the compatibility conditions between metric and connection. As it was, using the makeshift technique described above to get the Newtonian result, he was amazed to find that the spatial metric is non-flat. Even today, almost all treatments of the Newtonian limit of general relativity are still based on this makeshift approach that employs only the metric tensor.

10. CONCLUSION

The moral of this story is that general relativity is primarily a theory of an affine connection on a four-dimensional manifold, which represents the inertio-gravitational field. The other important spacetime structure is the metric field that represents the chrono-geometry; and the peculiarity of general relativity is that the compatibility conditions between metric and connection—or in physical terms, between inertio-gravitational field and chrono-geometry—uniquely determine the connection in terms of the metric. In teaching the subject, emphasis should be put on the connection from the beginning. This can be done easily by presenting the affine version of Newtonian gravitation theory before discussing general relativity. But most textbooks still start from the metric and introduce the connection later via the Christoffel symbols in a way that does not stress the basic role of the connection.[36] Now that gauge fields have come to dominate quantum field theory, it is more important than ever to emphasize from the beginning how general relativity resembles these Yang-Mills type theories, as well as how it differs.[37]

---

35 For details, see (Stachel 1989; Norton 1984) and volume 1 of this series.

36 It is indicative of current interests that (Darling 1994), the only elementary mathematical textbook I know that introduces the connection first, does not even mention the application to gravitation theory, but concludes with a chapter on "Applications to Gauge Field Theory" (pp. 223–250).

37 The basic difference is that the affine connection lives in the frame bundle (see Section h of the Appendix), which is soldered to the spacetime manifold. The symmetries of the fibres are thus induced by spacetime diffeomorphisms. On the other hand, the Yang-Mills connections live in fibre bundles, the fibres of which have symmetry groups that are independent of the spacetime symmetries (internal symmetries). For further discussion, see (Stachel 2005).

ACKNOWLEDGEMENTS AND A CRITICAL COMMENT

I thank Dr. Jürgen Renn for a thoughtful reading of this paper, and many helpful suggestions for its improvement. I thank Dr. Erhard Scholz for his careful critique of the paper. While agreeing with its basic viewpoint, he made some critical comments on my treatment of Grassmann and the mythical Weylmann. With his kind permission I quote them:

> The (historical) *lineale Ausdehnungslehre* was so much oriented towards the investigations of linear geometric structures and their algebraic generalization that there was a deep conceptual gulf between Grassmann's approach and Riemann's differential geometry of manifolds, which could only be bridged after a tremendous amount of deep and hard work. I do not see in Grassmann's late attempt to understand the algebraic geometry of curves and surfaces in terms of his *Ausdehnungslehre* a step that might have led him even somewhat near to a generalization of parallel transport in the sense of differential geometry. In "real history" there was no natural candidate for "Weylmann."
>
> ..... So, in short, your Newstein paper is an interesting thought experiment discussing the question of what would have happened if history had gone other than it did. In doing so, and following your line of investigation, we might find more precise answers as to why there was, e.g., still a long way to go from Grassmann to a potential "Weylmann." This is contrary to your intentions, I fear, but I cannot help reading your paper that way.

Rather than going contrary to my intentions, his remarks raise a most important question that supplements my approach to alternate histories: Given that we can invent various alternatives to the actual course of events, can one attach a sort of intrinsic probability to these various alternatives? I mean probability in the sense of a qualitative ranking of the probability of the alternatives rather than attaching a numerical value to the probability of each. In a truly "postmature" case, the ranking of the actual course of events would be lower than that of at least one of the alternatives. For example, the probability of a direct mathematical route from Riemann's local metric to Levi-Civita's local metrical parallelism would rank higher than the probability of the actual route via physics through Einstein's development of general relativity. Dr. Scholz makes a strong case for ranking the probability of the actual course of events from Grassmann's affine spaces to Weyl's affine connection higher than the probability of the step from Grassmann to Weylmann in my myth. I shall not pursue this issue further here, but again thank Dr. Scholz for comments that raise it in the context of my paper.

APPENDIX: RIEMANNIAN PARALLELISM AND AFFINELY CONNECTED SPACES

I shall review the concepts of parallelism in Euclidean and affine spaces, and their generalization to non-flat Riemannian and affinely connected spaces, respectively. I shall emphasize material needed to understand the historical and mathematical discussion in the Sections 3–6 and Newstein's mythical history in Section 7. Those familiar with the mathematical concepts may refer to the Appendix as needed when reading Sections 4–7.[38]

a. affine and Euclidean spaces. The familiar concept of parallelism in Euclidean space can easily be extended from lines to vectors: two vectors at different points in that space are parallel if they are tangent to parallel lines. We say that two Euclidean vectors are equal if they are parallel and have the same length as defined by the metric of Euclidean space. But, as we shall soon see, the concepts of parallelism and equality of parallel vectors retain their significance when we abstract from the metric properties of Euclidean space to get an affine space.

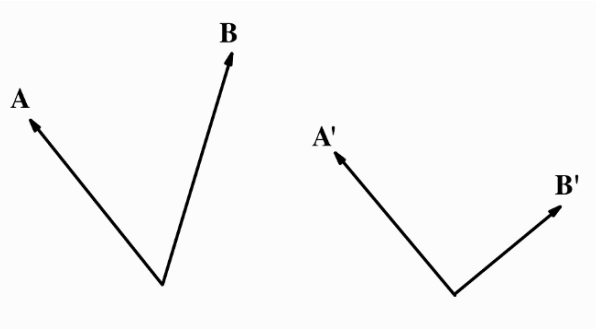

Figure 1: Any pair of non-parallel vectors $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *can be transformed into any other pair* $\boldsymbol{A}'$ *and* $\boldsymbol{B}'$ *by an (active) affine transformation.*

The properties of Euclidean geometry may be defined as those that remain invariant under transformations of the Euclidean group, consisting of translations $T(3, R)$,[39] and of rotations $O(3, R)$ about any point in space.[40] A translation is a point transformation that takes the point P into the point $P + \boldsymbol{v}$, where $\boldsymbol{v}$ is any vector. A rotation is a point transformation with a fixed point P that takes the point

38 However, in contrast to more familiar treatments, I shall define connections in terms of frame bundles, a concept that I shall introduce informally, following (Crampin and Pirani 1986, chaps. 13–15), which may be consulted for more details.

39 I shall use the notation (n, R) to denote a group acting on a real n- dimensional space.

40 I shall give the active interpretation of all geometrical transformations: The transformations act on the points of the space in question, taking each point into another one. The idea of defining a geometry by the group of transformations that leave invariant all geometric relations goes back to (Klein 1872).

$P + \boldsymbol{r}$ into the point $P + \boldsymbol{\mathcal{O}}\boldsymbol{r}$ where $\boldsymbol{\mathcal{O}} \in O(3, R)$ is an orthogonal transformation. The translations are clearly metric-independent; but the orthogonal transformations, being the linear transformations that preserve the distance between any pair of points, clearly do depend on the metric.

If we relax the condition that a linear transformation $\boldsymbol{L}$ preserve distances, and merely demand that it have a non-vanishing determinant), then $\boldsymbol{L} \in GL(3, R)$, the group of general linear or affine transformations. Together with the translations, they form the affine group that defines an affine geometry.[41] Parallelism of lines and vectors and the ratio of the lengths of parallel vectors (and hence the equality of two such vectors) being invariant under the affine group, are meaningful affine concepts. The (Euclidean) length of any vector is changed by an affine transformation with non-unit determinant, so it is not a meaningful affine concept.

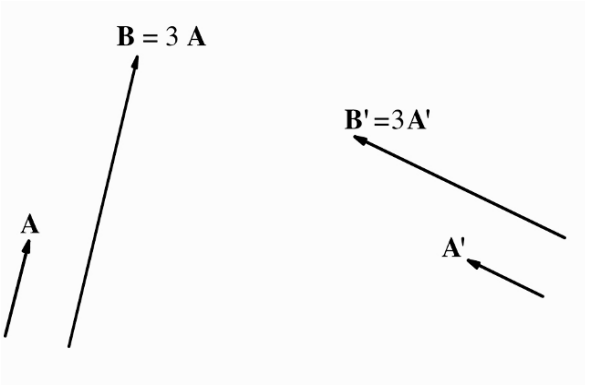

Figure 2: Any pair of parallel vectors $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *can be transformed into any other pair of parallel vectors* $\boldsymbol{A'}$ *and* $\boldsymbol{B'}$ *with the same ratio by an (active) affine transformation.*

In order to determine the action of an affine transformation $\boldsymbol{L}$ on any vector $\boldsymbol{v}$ at some point of an n-dimensional affine space, we need merely define its action on a basis or linear frame $\boldsymbol{e}_i$ at that point, consisting of n linearly-independent vectors:

$$\boldsymbol{e}_j{}' = \boldsymbol{L}^i_j \boldsymbol{e}_i, \tag{11}$$

where $\boldsymbol{e}_j{}'$ is the new basis produced by the action of $\boldsymbol{L}$ on $\boldsymbol{e}_i$, and $\boldsymbol{L}^i_j$ is the matrix representing the action of $\boldsymbol{L}$ on some basis. (Here and throughout, we have adopted the summation convention for repeated indices, which range over the appropriate number of dimensions—here $1, \ldots, n$.)

If we want to restrict ourselves to Euclidean geometry and the orthogonal group, we may restrict ourselves to orthonormal bases or frames:

41 For a discussion of affine and metric spaces, with a view to the generalizations needed below, see (Crampin and Pirani 1986, chaps. 1 and 7). For these generalizations, see chaps. 9 and 11.

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij}, \tag{12}$$

where the dot symbolizes the Euclidean scalar product of two vectors, and to orthogonal changes of bases:

$$\boldsymbol{e}_j{}' = \mathcal{O}^i_j \boldsymbol{e}_i, \qquad \boldsymbol{e}_i{}' \cdot \boldsymbol{e}_j{}' = \delta_{ij}. \tag{13}$$

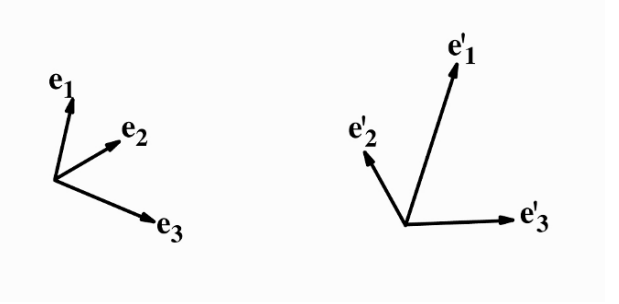

Figure 3: A (homogenous) affine transformation is defined by its action on a basis (or linear frame) $\boldsymbol{e}_A$ of the affine space.

Once we have chosen a basis at one point of an affine (or Euclidean) space, we can take as the basis at any other point of space the set of basis vectors equal and parallel to the original basis, thereby setting up a field of bases or *linear frames* over the entire space.

b. frame bundles. On the other hand, we can consider the set of all possible bases or linear frames at a given point of space. As is clear from eq. (11), in an affine space these frames are related to each other by the transformations of $GL(n, R)$. The set of all frames, together with the structure that the n-dimensional affine group imposes on them, is said to form a *fibre* over the point in question. Similarly, in Euclidean space, the set of all possible orthonormal frames at a point has a structure imposed on it by $O(n, R)$, the n-dimensional orthogonal group (see eq. (13)).

The set of all possible frames at every point of a space together with the space itself form a manifold that is called the *bundle of linear frames* or, more simply, the *frame bundle*. This is a special case of the more general concept of a *fibre bundle*.[42] The original space, which is affine or Euclidean in our examples but capable of generalization to any manifold, is called the *base space* of the fibre bundle; each *fibre* also need not be composed of linear frames, but may have a more general structure (below we shall consider fibres composed of tangent spaces). But there is always a *projection operation* that takes us from any fibre of the bundle to the point of the base

42 For fibre bundles in general and the frame bundle in particular, see (Crampin and Pirani 1986, chaps. 13 and 14).

space at which the fibre is located. A fibre bundle is called *trivial* if it is equivalent to the Cartesian product of a base manifold times a single fibre with a structure on it. The frame bundles we have been considering are trivial, since they are equivalent to the product of an affine (or Euclidean) space times a frame fibre with the structure imposed on it by the affine (or orthogonal) group.

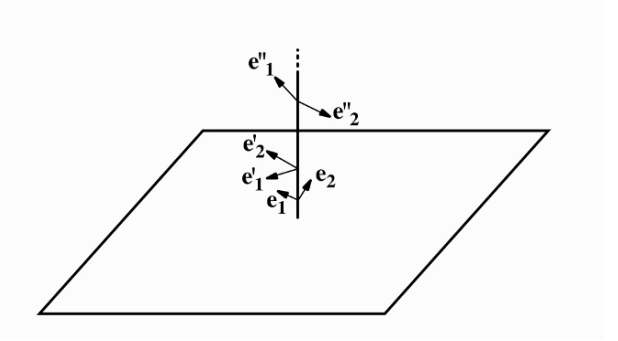

Figure 4: The set of all possible frames $\boldsymbol{e}_A$, $\boldsymbol{e}'_A$, $\boldsymbol{e}''_A$, … *at a point of the space forms a "fibre" over the point.*

A *cross-section* of the frame bundle is a specification of a particular frame on each fibre of the bundle, i.e., at each point of the base space (see Fig. 6). (The frames must vary in a smooth way as we pass from point to point, but we shall not bother here with such mathematical details.) In an affine (or Euclidean) space, the specification of a linear (or orthonormal) frame on one fibre allows us to pick out a unique parallel cross section of the entire bundle. (The last sentence just repeats, in the language of fibre bundles, something said earlier.) A change of frame on one fibre produces a change of the entire parallel cross section that is induced by an affine (orthogonal) transformation on the original fibre.

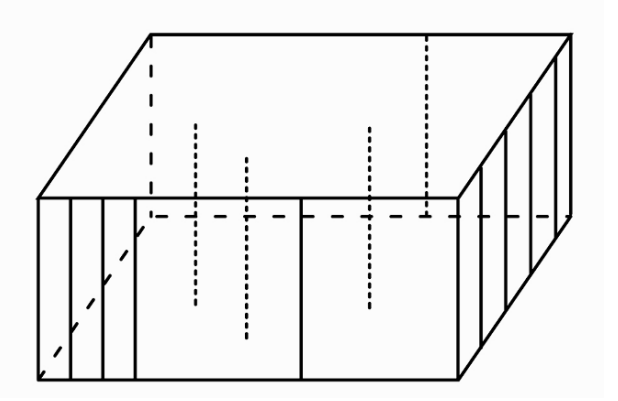

Figure 5: The fibres of a fibre bundle.

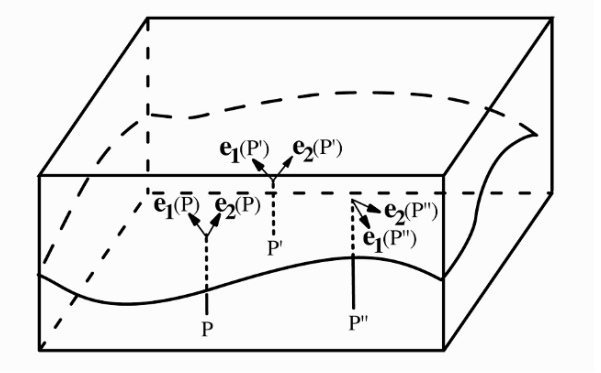

Figure 6: A "cross-section" of a frame bundle is a choice of a particular frame on each fibre of the bundle.

c. parallelism in non-flat Riemannian spaces. Now consider three-dimensional Euclidean space and some two-dimensional (generally curved) surface S in it. All vectors that are tangent to S at one of its points P form a vector space $T(P)$, called the tangent space to S at P. The collection of all such tangent spaces for all points $P \in S$ form a fibre bundle $T(S)$, called the *tangent bundle*. All vectors in $T(S)$ are intrinsically related to S,[43] and we want to define the concept of parallelism for such vectors in such a way that it will also be intrinsic to S. We cannot simply take the vector at another point P' of S that is parallel to a vector of $T(P)$ in the three-dimensional Euclidean sense: in general, that vector will not even be in $T(P')$, see Fig. 7).

We can get an idea of how to proceed by considering the case when S is a plane. The concept of parallel vectors at different points of the plane is clearly intrinsic to the plane. Consequently, the tangent spaces at each point of the plane can be identified with each other in a natural way, as can pairs of orthonormal vectors $\boldsymbol{e}_A (A = 1, 2)$ that form a basis at each point of the plane considered as a two-dimensional Euclidean space. Taken together with the unit normal vector $\boldsymbol{n}$ to the plane, the $\boldsymbol{e}_A$ form a basis for the tangent space of the three-dimensional Euclidean space.

43 These vectors can, for example, be defined as the tangent vectors to curves $C = P(s)$ lying entirely in S. We follow the usual terminology in distinguishing *curves* from *paths*, which are curves without a parametrization s.

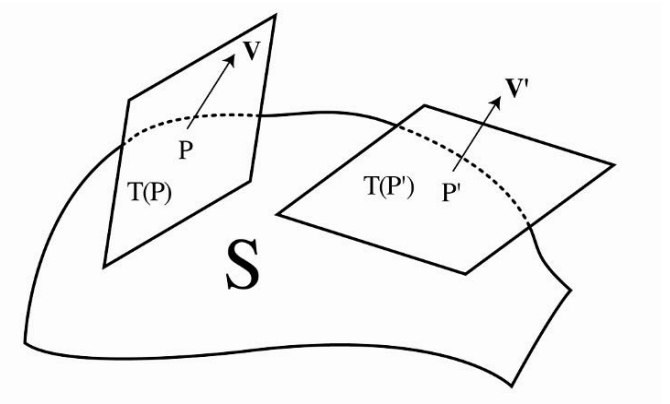

Figure 7: To define an intrinsic notion of parallelism within a surface S, we cannot use vectors that are parallel to each other in the three-dimensional sense. While V lies in the tangent plane at P, the three-dimensionally parallel vector V' does not even lie in the tangent plane at P'.

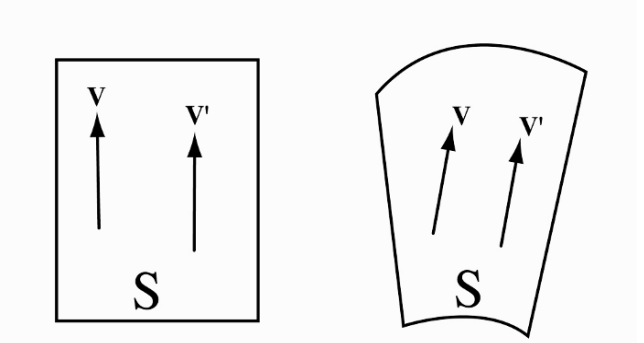

Figure 8: If V and V' are parallel vectors in the plane S, and if parallelism is intrinsic to S, then they remain parallel even when S is bent (without distortion).

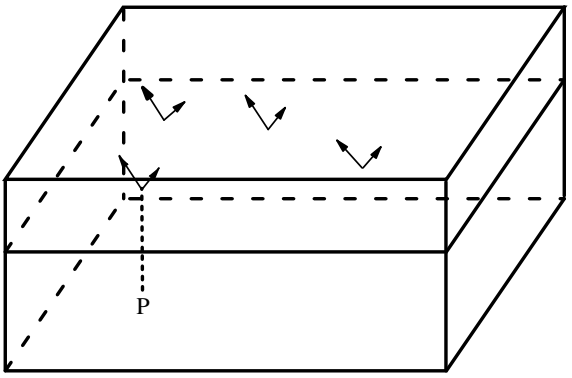

Figure 9: In an affine space, choice of a frame on one fibre picks out a unique parallel cross-section.

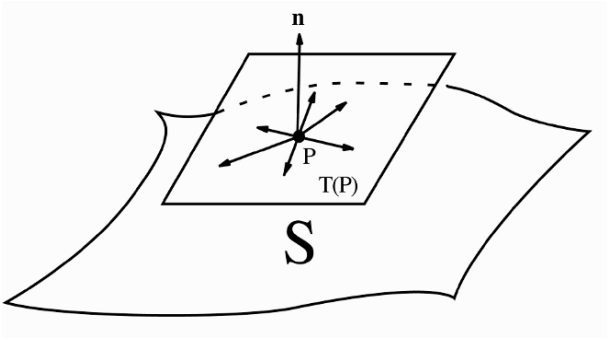

Figure 10: The tangent space $T(P)$ to a surface S at point P of the surface in Euclidean space is composed of all vectors tangent to the surface at that point. The unit normal to the tangent plane is designated by $\boldsymbol{n}$.

Now suppose we bend the plane without distorting its metric properties (i.e., the metrical relations between its points as measured *on the surface*), resulting in what is called a *developable surface*.[44] If we want the concepts of parallelism and straight line to be intrinsic to a such a surface, they must remain the same for any surface developed from the plane as they were for the plane itself. Thus, the basis vectors $\boldsymbol{e}_A$

44 Such a process of bending leaves the *intrinsic* geometry of the surface unchanged, but changes its *extrinsic* geometry. The intrinsic properties of any surface are those that remain unchanged by all such bendings; its extrinsic properties are precisely those that depend on how the surface is embedded in the enveloping Euclidean space.

at different points of the surface must still be considered parallel to each other from the intrinsic, surface viewpoint, even though they are not from the three-dimensional Euclidean point of view. Consider two neighboring points on the surface P and $P' = P + d\boldsymbol{r}$. In order to get from the tangent plane $T(P)$ at P to the tangent plane $T(P')$ at P' one must rotate the former through the angle $d\theta$ that takes $\boldsymbol{n}$ into $\boldsymbol{n}'$.[45] Thus, there must be an orthogonal transformation $\boldsymbol{\mathcal{O}}$, differing from the identity $\boldsymbol{I}$ only by an amount that depends on P, P', $\boldsymbol{n}$ and $\boldsymbol{n}'$, or equivalently on P, $\boldsymbol{n}$, $\boldsymbol{n}'$ and $d\boldsymbol{r}$:

$$\boldsymbol{\mathcal{O}} = \boldsymbol{I} + d\boldsymbol{\mathcal{O}}, \quad d\boldsymbol{\mathcal{O}} = d\boldsymbol{\mathcal{O}}(P, \boldsymbol{n}, \boldsymbol{n}', d\boldsymbol{r}) \tag{14}$$

and depends linearly on $d\boldsymbol{r}$.

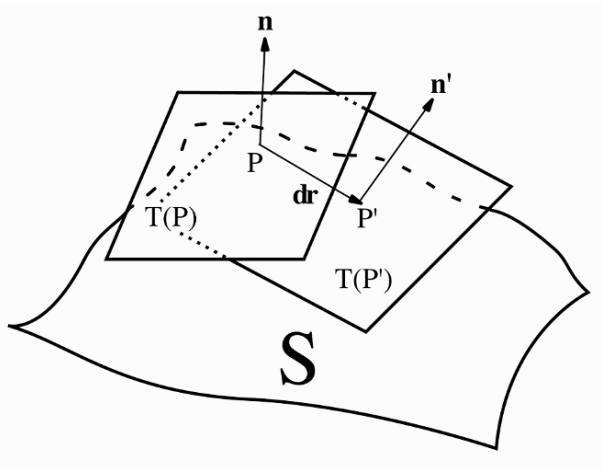

Figure 11: In order to get from the tangent plane $T(P)$ at P to a neighboring tangent plane $T(P')$ at $P' = P + d\boldsymbol{r}$, we must carry out an orthogonal transformation $\boldsymbol{\mathcal{O}} = \boldsymbol{I} + d\boldsymbol{\mathcal{O}}$ that depends on P, $d\boldsymbol{r}$, $\boldsymbol{n}$ and $\boldsymbol{n}'$.

Due to the linearity of vector spaces, the effect of this orthogonal transformation on any vector in the tangent plane to the surface can be computed once its effect on a set of basis vectors $\boldsymbol{e}_A$ in the tangent plane is known.[46] The change in each basis vector is given by:

$$\delta\boldsymbol{e}_B(P') = (d\boldsymbol{O})^A_{\ B}\, \boldsymbol{e}_A, \tag{15}$$

45 The concept of parallelism in the Euclidean space allows us to draw the vector at P that is equal and parallel to $\boldsymbol{n}'$ at P', and so define the angle $d\theta$ between $\boldsymbol{n}$ and $\boldsymbol{n}'$.

46 Note that we need the normals $\boldsymbol{n}$ and $\boldsymbol{n}'$ to define the orthogonal transformation between parallel vectors lying in the tangent planes at P and P'; but since we are only interested in the change in vectors lying in the surface we may omit $\boldsymbol{n}$ from explicit mention in eq. (5), since it is determined by the $\boldsymbol{e}_A$ and the orthonormality conditions.

where $(d\mathbf{0})^A_B\ (P, d\boldsymbol{r})$ are the elements of a matrix that determines the effect of the infinitesimal rotation on the orthonormal basis vectors.

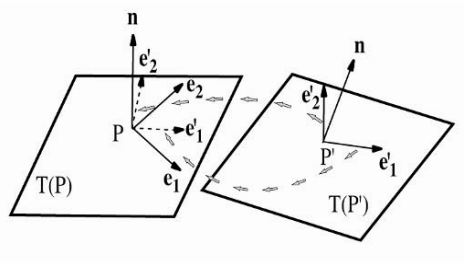

Figure 12: The effect of $d\boldsymbol{O}$ on any vector in $T(P)$ is determined by its effect on a set of basis vectors $\boldsymbol{e}_A$ of the space.

It is this connection between parallel vectors in neighboring tangent planes, given by eqs. (14) and (15), that we shall preserve for all surfaces, in particular for those that are not intrinsically plane. Since it was introduced by Levi-Civita (see Section 4), it is often called the Levi-Civita connection. If two points P and Q are not neighboring, we must choose some path C on the surface connecting P and Q, and break it up into small straight line segments $PP', P'P'', ..., Q$. If we move from P to P' along straight-line segment PP', we must rotate $T(P)$ at P through some small angle $\Delta\theta$ about the normal $\boldsymbol{n}$ at P in order to get the tangent plane $T(P')$ at P'. For the next segment $P'P''$, we have to rotate the tangent plane $T(P')$ through an angle $\Delta\theta'$ about the normal $\boldsymbol{n}'$ at P' in order to get the tangent plane $T(P'')$ at P''. We keep doing this until we reach the endpoint Q. Now we increase the number of intermediate points indefinitely, and take the limit of this process so that the broken straight line segments approach the curve. This defines the vector in $T(Q)$ that is parallel to one in $T(P)$ *with respect to the path* C.

Note that we must add the last qualification because, unless S is a developable surface, the resulting parallelism in general will be path dependent. We can see this by looking at a small parallelogram with sides $PP', P'Q$ and $PP'', P''Q$. Since $T(P')$ and $T(P'')$ are not in general parallel to each other, the correspondence between vectors in the tangent planes $T(P)$ and $T(Q)$ that is set up by going via $T(P')$ is not in general the same as the one we get by going via $T(P'')$.

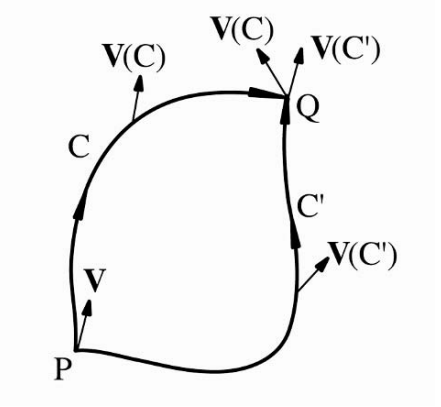

Figure 13: In general the vector $\mathbf{V}'(C)$ at Q that is parallel to $\mathbf{V}$ at P depends on the path taken between P and Q.

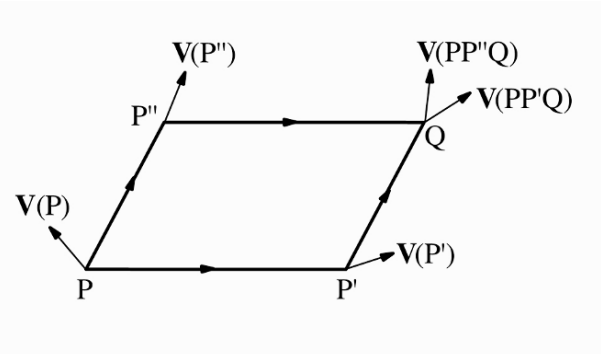

Figure 14: We can see this by looking at the parallel transport of a vector $\boldsymbol{v}$ along the sides PP', $P'Q$ and PP'', $P''Q$ of a small parallelogram.

d. the Riemann tensor. By carrying out the analysis of this parallelogram quantitatively, we can define the Riemann tensor of the surface.[47] Take a vector $\boldsymbol{v}$ in $T(P)$, and let the corresponding (i.e. intrinsically parallel) vector in $T(P')$ be $\boldsymbol{v} + \delta\boldsymbol{v}$. Then $\boldsymbol{v} + \delta\boldsymbol{v}$ results from $\boldsymbol{v}$ by a rotation operation that acts on $\boldsymbol{v}$; we shall symbolize it by the operator $\boldsymbol{I} + d\boldsymbol{\Omega}$ (see eq. (14)), so that:

47 In the case of a two-dimensional surface, it reduces to a scalar R; i.e., all non-vanishing components of the Riemann tensor reduce to $\pm R$. But we prefer to keep the tensorial designation in view of the impending generalization to higher dimensions.

$$\delta \boldsymbol{v} = d\boldsymbol{\mathcal{O}}\boldsymbol{v}.$$

Here, $d\boldsymbol{\mathcal{O}}$ represents a first order infinitesimal rotation operator that depends linearly on $d\boldsymbol{r}$. Similarly, if $d\boldsymbol{r}'$ represents the displacement PP'', then the change $\delta\boldsymbol{v}'$ in $\boldsymbol{v}$ when we go from $T(P)$ to $T(P'')$ is given by:

$$\delta \boldsymbol{v}' = d\boldsymbol{\mathcal{O}}'\boldsymbol{v}.$$

Then the change in $\boldsymbol{v}$ at $T(Q)$ when we go via $d\boldsymbol{r}$ first, then $d\boldsymbol{r}'$ (i.e., via $PP'Q$) is given by:

$$\delta \boldsymbol{v}_1 = (\mathcal{I} + d\boldsymbol{\mathcal{O}}')(\mathcal{I} + d\boldsymbol{\mathcal{O}})\boldsymbol{v} - \boldsymbol{v} = (d\boldsymbol{\mathcal{O}}' + d\boldsymbol{\mathcal{O}} + d\boldsymbol{\mathcal{O}}' d\boldsymbol{\mathcal{O}} v)\boldsymbol{v};$$

while, if we proceed in the reverse order (i.e., via $PP''Q$), the change is given by:

$$\delta \boldsymbol{v}_2 = (d\boldsymbol{\mathcal{O}} + d\boldsymbol{\mathcal{O}}' + d\boldsymbol{\mathcal{O}} d\boldsymbol{\mathcal{O}}')\boldsymbol{v}.$$

Since $\delta\boldsymbol{v}_1$ and $\delta\boldsymbol{v}_2$ are vectors at the same point, their difference is a (second order infinitesimal) vector $\delta^2\boldsymbol{v}$. It indicates by how much the two vectors in $T(Q)$ that are parallel to $\boldsymbol{v}$ in $T(P)$, depending on which of the two paths is taken, differ from each other:

$$\delta^2 \boldsymbol{v} = (d\boldsymbol{\mathcal{O}} d\boldsymbol{\mathcal{O}}' - d\boldsymbol{\mathcal{O}}' d\boldsymbol{\mathcal{O}})\boldsymbol{v}.$$

Note the operator in parentheses is the same for all vectors in $T(P)$ since they are all rotated by the same amount. And since $d\boldsymbol{\mathcal{O}}$ and $d\boldsymbol{\mathcal{O}}'$ are linear in $d\boldsymbol{r}, d\boldsymbol{r}'$, respectively, this operator is proportional to $(d\boldsymbol{r} d\boldsymbol{r}' - d\boldsymbol{r}' d\boldsymbol{r})$. Such an antisymmetric tensorial product of two vectors is abbreviated as $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ and called a simple bivector; it represents the (signed) area of the infinitesimal parallelogram with sides $d\boldsymbol{r}, d\boldsymbol{r}'$. This second order infinitesimal term is also same for all vectors taken from P to Q along the sides of the parallelogram.[48] So there must be a finite tensorial operator $\boldsymbol{R}$, such that, when it operates on an area bivector $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ and a vector $\boldsymbol{v}$, it produces the change in $\boldsymbol{v}$ when it is parallel transported around the area $d\boldsymbol{r} \wedge d\boldsymbol{r}'$. Note that, to the second differential order we are considering, it makes no difference whether we parallel transport a vector from P to Q in two different ways, and compare the results in $T(Q)$, or take it around the parallelogram and compare the result with the original vector in $T(P)$. Further, the result is independent of the shape of the infinitesimal plane figure we carry it around so long as this has the same area as, and lies in the plane defined by, $d\boldsymbol{r} \wedge d\boldsymbol{r}'$. The tensorial operator $\boldsymbol{R}$, which operates on a bivector and a vector to produce another vector, is called the *Riemann tensor*; when it operates on an infinitesimal area element, it measures how much Riemannian parallelism

48 One should actually distinguish between $d\boldsymbol{r}$ at P and $d\boldsymbol{r}$ at P'', which is the result of parallel transporting $d\boldsymbol{r}$ at P along $d\boldsymbol{r}'$. But to the order we are considering, the difference may be neglected. The more serious problem of whether the parallelogram resulting from these displacements actually "closes" will be discussed later.

on that surface element differs from flat, path-independent, parallelism, for which the Riemann tensor would vanish.

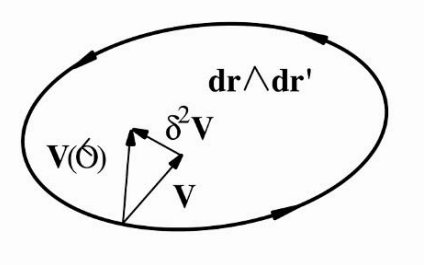

Figure 15: The operator $\boldsymbol{R}$, *operating on the area* $d\boldsymbol{r} \wedge d\boldsymbol{r}'$ *and the vector* $\boldsymbol{v}$, *produces the change* $\delta^2\boldsymbol{v}$ *in* $\boldsymbol{v}$ *when it is parallel transported around that area.*

e. non-flat affine spaces. Our discussions of parallelism on a surface and of the Riemann tensor made essential use of the metric of the enveloping Euclidean space. First of all, this metric induced a notion of distance on the surface; but this is intrinsic to the surface, and can be defined without using the fact that the surface is embedded in a Euclidean space. More serious is the fact that we used the normals to the surface at each point in order to develop the relation between tangent spaces at neighboring points in terms of an orthogonal transformation (rotation through some angle). The notions of orthogonality and angle are intrinsically metrical.

Suppose we abstract from these metric concepts and consider an affine space, as discussed above. Using only affine concepts, can we still define concepts of parallelism and straight line on a surface in an affine space? The answer is yes, but we must introduce a substitute for the unit normal field given naturally in a Euclidean space.

First of all, the concept of surface is independent of a metric, as are those of the tangent space at each point of a surface, and (hence) of the tangent bundle. But now we have no natural way of relating the tangent spaces at different point of the surface by means of a general linear transformation. At each point of the surface, a basis in its tangent space must be supplemented by a vector that does not lie in the tangent space; i.e., a vector that takes the place of the normal vector to the surface in a Euclidean space. Together with the chosen basis in the tangent space, this vector constitutes a basis for the enveloping affine space. This vector field is said to *rig* the surface, and the process is called *rigging*. Once the surface is rigged, one can carry out in an affine space a procedure to relate neighboring tangent spaces that is entirely analogous to the procedure used in the Euclidean case. The only difference is that, instead of the infinitesimal orthogonal transformation $d\mathcal{O}$ that carries the orthonormal basis at P into the orthonormal basis at P', one considers the infinitesimal general linear transformation $d\mathcal{L}$ that takes a basis for the enveloping affine space at P into the corresponding basis at P'. Due to the linearity of vector spaces, carrying out the transformation $d\mathcal{L}$ on any vector in $T(P)$ yields the corresponding parallel vector in $T(P')$. Such a connection between tangent spaces, which generalizes to surfaces in

an affine space the Levi-Civita connection for surfaces in a Euclidean space, is called *a general linear connection*. Once the connection is defined, everything proceeds in a way that is entirely analogous to that for Euclidean spaces (see the previous subsection), up to and including the definition of the Riemann tensor operator.

Instead of eq. (15), giving the effect of an infinitesimal orthogonal transformation (rotation) matrix on an orthonormal basis, we can now specify the effect on the vectors in a tangent plane of an infinitesimal general linear transformation, by specifying the infinitesimal general linear transformation matrix $(dL)^A_B$ that gives the effect of this transformation on an arbitrary basis:

$$\delta \boldsymbol{e}_B(P') = (dL)^A_B \boldsymbol{e}_A. \tag{16}$$

f. covariant differentiation, geodesics. Once we have the concept of parallelism along a path, we can define a derivative operation for a vector field on a surface. The essence of the usual derivative operation for a vector field in Euclidean space consists in comparing the value of the vector field $\boldsymbol{v}$ at some point with its values at some neighboring points. But we can only compare vectors in the same tangent space: what we actually do to compare vectors at two points P and Q is to compare $v(Q)$ with the vector at Q that is parallel to $\boldsymbol{v}(P)$. We shall proceed in the same way on a surface and compare values at two neighboring points P and $P + d\boldsymbol{r}$:

$$\boldsymbol{v}(P + d\boldsymbol{r}) - [\boldsymbol{v}(P) + \delta \boldsymbol{v}] = d\boldsymbol{r} \cdot \partial \boldsymbol{v} - \delta \boldsymbol{v},$$

since

$$\boldsymbol{v}(P + d\boldsymbol{r}) = \boldsymbol{v}(P) + d\boldsymbol{r} \cdot \partial \boldsymbol{v},$$

where ∂ represents the ordinary-derivative gradient operation; operating on a scalar field $\phi(\boldsymbol{r})$, it gives the gradient vector field $\partial\phi$; but operating on a vector (or tensor) field it does not produce another vector (or tensor) field. It must be supplemented by the second term $\delta \boldsymbol{v}$ for a vector (and similar terms for higher-order tensors). Since $\delta \boldsymbol{v} = d\boldsymbol{\mathcal{O}}\boldsymbol{v}$, we can write the invariant combination as

$$\boldsymbol{v}(P + d\boldsymbol{r}) - [\boldsymbol{v}(P) + \delta \boldsymbol{v}] = d\boldsymbol{r} \cdot \partial \boldsymbol{v} - d\boldsymbol{\mathcal{O}}\boldsymbol{v}.$$

Since $d\boldsymbol{\mathcal{O}}\boldsymbol{v}$ is also linear in $d\boldsymbol{r}$, we can abbreviate the right-hand side as:

$$d\boldsymbol{r} \cdot \partial \boldsymbol{v} - d\boldsymbol{\mathcal{O}}\boldsymbol{v} = d\boldsymbol{r} \cdot \nabla \boldsymbol{v}.$$

The expression $d\boldsymbol{r} \cdot \nabla$ represents an invariant directional derivative in the $d\boldsymbol{r}$ direction. Since the result is linear in $d\boldsymbol{r}$, there must be a tensorial operator $\boldsymbol{\nabla}$ called the *covariant derivative* operator, that operates on a vector to produce a mixed tensor $\boldsymbol{\nabla v}$ with one covariant and one contravariant (i.e.,vectorial) place.

On a surface, we may generalize the concept of a straight line in an affine space to that of a geodesic by requiring that the parallel transport of its tangent vector along a geodesic remain the tangent vector. If $\boldsymbol{t}$ represents the tangent vector to the curve $C(\lambda)$, $\boldsymbol{t} = dC/d\lambda$, this means that a geodesic must satisfy the equation:

$$\boldsymbol{t} \cdot \nabla \boldsymbol{t} = 0.$$

g. generalizations, intrinsic characterizations. Nothing in the discussion above depends essentially on the number of dimensions being three, and it can be immediately generalized to n-dimensional metric and affine spaces, defined by the translation groups $T(n)$ and $O(n, R)$ and $T(n)$ and $GL(n, R)$ respectively; and to their m-dimensional sub-spaces. If m is less than $n-1$, then there are n-m normals, and n-m rigging vectors must be defined; but otherwise the discussion proceeds quite analogously. Since any m-dimensional Riemannian or affinely-connected space can be embedded in an n-dimensional Euclidean or affine space of sufficiently high dimension (locally, if not globally), such embedding arguments can handle the generic case.

Of course, once the basic geometrical concepts have been grasped, an intrinsic method of characterizing curved spaces, independently of any embedding in flat spaces of higher dimension, is preferable. It is clear from the previous discussion how to proceed. One must specify a connection between vectors in $T(P)$ and $T(P')$ that defines when a vector in one is parallel to a vector in the other. In contrast to the order in the previous embedding considerations, I shall first give the definition for a general affine linear connection, and then indicate how to specialize it to a Riemannian or Levi-Civita connection.

As indicated earlier (see discussion around eqs. (15) and (16) above), in order to connect arbitrary vectors in the two tangent spaces, it suffices to indicate how sets of basis vectors in the two tangent spaces are connected. Let $\boldsymbol{e}_i(P)$ be a set of basis vectors in $T(P)(i = 1, 2, \ldots, n)$. The changes in these basis vectors when we move to $T(P + d\boldsymbol{r})$ will be given by (generalizing eq. (6) above):

$$\delta \boldsymbol{e}_i(P') = (dL)_i^j \, \boldsymbol{e}_j, \qquad (i, j = 1, 2, \ldots, n).$$

Our connection is linear in $d\boldsymbol{r}$, so it suffices to know the change in $\boldsymbol{e}_i$ for a small change in each of the basis directions, $d\boldsymbol{r} = \epsilon e_k$, where ϵ is an infinitesimal of first order.

$$\delta \boldsymbol{e}_i(P') = \epsilon L_i^j(P, \boldsymbol{e}_k) \boldsymbol{e}_j.$$

On the other hand $\delta \boldsymbol{e}_i$ itself must be a linear combination of the basis vectors, so we may decompose it into the infinitesimal changes in each of these directions:

$$(\delta \boldsymbol{e}_i)_j = \epsilon \Gamma_{ij}^k \boldsymbol{e}_k.$$

Thus, specification of the set of quantities $\Gamma_{ij}^k(P)$ at all points of the manifold fixes the affine connection intrinsically.[49] We call the Γ_{ij}^k the *components of the connection* with respect to the basis $\boldsymbol{e}_i$.[50]

If we now want to construct the parallelogram as described above in the definition of the Riemann tensor, we must make sure that it "closes," that is, that we reach the same point if we parallel transport $d\boldsymbol{r}$ along $d\boldsymbol{r}'$ as we do if we parallel transport $d\boldsymbol{r}'$

along $d\boldsymbol{r}$. It is relatively simple to show that this will be the case if Γ^k_{ij} is symmetric in its two lower indices; we shall consider only such symmetric affine connections.[51]

The Riemann tensor operator $\boldsymbol{R}$ can now be defined in terms of its effect on the basis vectors. If we transport $\boldsymbol{e}_k$ around an area defined by $\boldsymbol{e}_i \wedge \boldsymbol{e}_j$, then its change in the $\boldsymbol{e}_l$ direction is given by R^k_{ijl}. These are the components of the Riemann tensor with respect to the basis $\boldsymbol{e}_i$, which can easily be related to the derivatives of the Γ^j_{ki}, but we omit the details. For future reference, we note that R^k_{ijl} is antisymmetric in its last pair of indices, and that if we contract its upper index with either of the last two indices, say the second, we get (plus or minus) the Ricci tensor R_{il}.

The covariant derivative operator will have components:

$$\nabla_i = \boldsymbol{e}_i \cdot \nabla;$$

the components of the covariant derivative of a vector $\boldsymbol{\nabla v}$, for example, are:

$$\nabla_i v^j = \partial_i v^j + \Gamma^j_{ki} v^k.$$

The components of the geodesic equation in an adapted coordinate system are:

$$d^2x^j/d\lambda^2 + \Gamma^j_{ki}(dx^k/d\lambda)(dx^i/d\lambda) = 0.$$

The components of the Riemann tensor with respect to a basis can be similarly calculated.

Turning to Riemannian spaces, it is natural to demand that parallel transport along any path preserve the length of all vectors. If we impose this condition on a symmetric affine connection, we are led uniquely to the Levi-Civita connection discussed above; but again we omit the details.

For future reference, we also note that, just as in the case of a surface in a linear (flat) affine space discussed above, a connection is induced on a hypersurface in a non-flat affinely connected space if that hypersurface is rigged with an arbitrary vector field.[52]

---

49 Note that these quantities transform as scalars under a coordinate transformation, but as tensors under a change of basis. If we use the natural basis associated with a coordinate system (see the following note) and carry out a simultaneous coordinate transformation and change of natural basis, they transform under a more complicated, non-tensorial transformation law (see Section 6, eq. (7)).

50 Note that a basis need not be holonomic, i.e., coordinate forming. It will be if and only if the Lie bracket of any pair of basis vectors vanishes. We shall only need holonomic bases, for which an associated coordinate system x^i exists, such that in this coordinate system e^j_i, the coordinate components of $\boldsymbol{e}_i$, are equal to δ^j_i, the Kronecker delta. Conversely, a basis is associated with any coordinate system by the same relations.

51 If the parallelogram does not close, the antisymmetric part of Γ^j_{ki} defines the so-called *torsion tensor*.

52 It is customary, when discussing spaces of more than three dimensions, to refer to subspaces of one less dimension than that of the space as *hypersurfaces*. Thus, when the discussion is generalized to more than three dimensions, "surfaces" become "hypersurfaces."

h. frame bundles and connections. We introduced the concept of affine connection in the currently-habitual way, in terms of its local action on vector or frame fields in some manifold. But a connection is more naturally introduced globally in terms of the frame bundle over that base manifold (see Section b). A curve C in the base manifold together with a frame field defined along the curve corresponds to a curve $\boldsymbol{C}$ in the frame bundle; and conversely $\boldsymbol{C}$ projects down to C in the base manifold, together with a frame field along the curve. Now a connection provides a rule for defining such curves in the frame bundle: given a curve C in the base manifold together with an initial frame at some point on the curve, parallel transport of the initial frame along the curve thus defines a unique curve $\boldsymbol{C}$ in the frame bundle. The only thing we have to worry about is what happens if we change the initial frame by the action of some element $\boldsymbol{L}^i_j$ of the general linear group (see eq. (11)).The curve in the frame bundle is then transformed into another curve that differs from the first only by the same action of $\boldsymbol{L}^i_j$ on the frame at each point of the curve in the base manifold.

We can use this idea to define a connection globally as a collection of curves in the frame bundle, each passing only once through any fibre of the bundle, that satisfy the following condition: if two such curves $\boldsymbol{C}$ and $\boldsymbol{C}'$ project into the same curve C in the base manifold, and hence have all of their fibres in common, then on each fibre the frames on the two curves are related by the global application of the same $\boldsymbol{L}^i_j$.

If we want to restrict the structure group of the frame fibres to some subgroup of $GL(n, r)$, then we must assure that the connection introduced is compatible with the structure of this subgroup. For example, if we required compatibility with any of the orthogonal or pseudo-orthogonal subgroups, the Levi-Civita connection would result.[53]

REFERENCES

Bhaskar, Roy. 1993. *Dialectic: The Pulse of Freedom*. London/New York: Verso.

Bonola, Roberto. 1955. *Non-Euclidean Geometry: A Critical and Historical Study of its Development*. Dover reprint: New York: Dover Publications. (Open Court 1912)

Cartan, Élie. 1923. "Sur les variétés à connection affine et la théorie de la relativité généralisée." *Ecole Normale Supérieure* (Paris). *Annales* 40: 325–412. English translation in (Cartan 1986).

———. 1986. *On Manifolds with an Affine Connection and the Theory of Relativity*. Translated by Anne Magnon and Abhay Ashtekar. Naples: Bibliopolis. (Chapter 1 printed in this volume.)

Collier, Andrew. 1994. *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*. London/New York: Verso.

Coolidge, Julian Lowell. 1940. *A History of Geometrical Methods*. Oxford: Clarendon Press.

Crowe, Michael J. 1994. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System*. New York: Dover Publications (reprint ed.).

Crampin, Michael, and Felix A.E. Pirani. 1986. *Applicable Differential Geometry*. Cambridge/London/New York: Cambridge University Press.

Darling, R. W. 1994. *Differential Forms and Connections*. Cambridge/New York/Melbourne: Cambridge University Press.

Ehlers, Jürgen. 1981. "Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie." In J. Nitsch et al. (eds.), *Grundlagenprobleme der modernen Physik*. Mannheim: Bibliographisches Institut, 65–84.

53 See (Crampin and Pirani 1986, chap. 15) for details.

Einstein, Albert. 1907. "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen." *Jahrbuch der Radioaktivität und Elektronik*. 4: 411–462.

———. 1914. "Die formale Grundlage der allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Mathematisch-physikalische Klasse*. *Sitzungsberichte*: 1030–1085.

———. 1915. "Zur allgemeinen Relativitätstheorie." *Königlich Preussische Akademie der Wissenschaften* (Berlin). *Mathematisch-physikalische Klasse*. *Sitzungsberichte*: 778–786.

Einstein, Albert and Marcel Grossmann. 1913. *Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation*. *I*. *Physikalischer Teil von Albert Einstein*. *II*. *Mathematischer Teil von Marcel Grossmann*. Leipzig: Teubner.

Eisenstaedt, Jean and A. J. Kox (eds.). 1992. *Studies in the History of General Relativity* (*Einstein Studies*, vol 3). Boston/Basel/Berlin: Birkhäuser.

Friedrichs, Kurt O. 1927. "Eine invariante Formulierung des Newtonschen Gravitationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz." *Mathematische Annalen* 98: 566–575.

Gauss, Carl Friedrich. 1902. *General Investigations of Curved Surfaces*. Princeton: Princeton University Press. English translation by A. Hiltebeitel and J. Morehead of *Disquisitiones generales circa superficies curvas* (Göttingen: Dietrich, 1828).

Grassmann, Hermann. 1844. *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*. Leipzig: Otto Wigand.

———. 1862. *Die Ausdehnungslehre*. *Vollständig und in strenger Form bearbeitet*. Berlin: Enslin.

———. 1877. "Ueber die Beziehung der nicht-Euklidischen Geometrie zur Ausdehnungslehre."Appendix I to (Grassmann 1878). English translation in (Grassmann 1995, 279–280).

———. 1878. *Die Ausdehnungslehre von 1844 oder Die lineale Ausdehnungslehre*. (2nd. Edition.) Leipzig: Otto Wigand.

———. 1995. *A New Branch of Mathematics: The Ausdehnungslehre of 1844 and Other Works*. Translated by Lloyd C. Kannenberg. Chicago and LaSalle: Open Court. (Appendix printed in this volume.)

Hertz, Heinrich. 1894. *Die Prinzipien der Mechanik*. *In neuem Zusammenhange dargestellt*, Philipp Lenard (ed.). *Gesammelte Werke*, Vol. 3. Leipzig: Barth.

Hessenberg, Gerhard. 1917. "Vektorielle Begründung der Differentialgeometrie." *Mathematische Annalen* 78: 187–217.

Howard, Don and John Stachel (eds.). 1989. *Einstein and the History of General Relativity* (*Einstein Studies*, Vol. 1). Boston/Basel/Berlin: Birkhäuser.

Janssen, Michel. 1992. "H. A. Lorentz's Attempt to Give a Coordinate-Free Formulation of the General Theory of Relativity." In (Eisenstaedt and Kox 1992, 344–363).

Jost, Jürgen. 1991. *Riemannian Geometry and Geometric Analysis*, 2nd. ed. Berlin/Heidelberg/New York: Springer.

Klein, Felix. 1872. *Vergleichende Betrachtungen über neuere geometrische Forschungen*. Erlangen: A. Dühechert. Revised version in *Mathematische Annalen* 43: 63–100 (1893).

Kolmogorov, Andrei N. and Adolf P. Yushkevich. 1996. *Mathematics of the 19th Century*. Basel/Boston/Berlin: Birkhäuser.

Laptev, B. L. and B. A. Rozenfel'd. 1996. "Chapter 1. Geometry." In (Kolmogorov and Yushkevich 1996, 1–118).

Lawvere, F. William. 1996. "Grassmann's Dialectics and Category Theory." In (Schubring 1996, 255–264).

Levi-Civita, Tullio. 1916. "Nozione de Parallelismo in una Varieta Qualunque e Conseguente Specificazione Geometrica della Curvatura Riemanniana." *Rendiconti del Circolo Matematico di Palermo* 42: 17–205. (English translation of excerpts given in this volume.)

Lützen, Jesper. 1995a. "Interactions Between Mechanics and Differential Geometry in the 19th Century." *Archive for History of Exact Sciences* 49: 1–72.

———. 1995b. "Renouncing Forces: Geometrizing Mechanics. Hertz's Principles of Mechanics." *Københavns Universitet Matematisk Institut, Preprint Series* 1995, No. 22.

Norton, John. 1984. "How Einstein Found his Field Equations: 1912–1915." *Historical Studies in the Physical Sciences* 14: 253–316.

Reich, Karen. 1992. "Levi -Civitasche Parallelverschiebung, affiner Zusammenhang, Übertragungsprinzip: 1916/17–1922/23." *Archive for History of Exact Sciences* 44: 78–105.

———. 1994. *Die Entwicklung des Tensorkalküls*. *Vom absoluten Differentialkalkuel zur Relativitätstheorie*. Basel: Birkhäuser.

Riemann, Bernhard. 1868. "Über die Hypothesen, welche die Geometrie zu Grunde liegen." In Richard Dedekind and H. Weber (eds.), *Gesammelte mathematische Werke*, 2nd ed. 1892, (reprinted in New York: Dover Publications, 1953), 272–287.

Scholz, Erhard. 1995. "Hermann Weyl's Purely Infinitesimal Geometry." In *Proceedings of the International Congress of Mathematicians, Zurich, Switzerland 1994*. Basel: Birkhäuser.

Schouten, Jan A. 1918. "Die direkte Analysis zur neueren Relativitätstheorie." *Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam*, XII, no. 6.

Schubring, Gert (ed.). 1996. *Hermann Guenther Grassmann (1809–1877): Visionary Mathematician, Scientist and Neohumanist Scholar*. Dordrecht/Boston/London: Kluwer Academic.

Stachel, John. 1986. "What a Physicist Can Learn from the Discovery of General Relativity." In Remo Ruffini (ed.), *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity*. Amsterdam: Elsevier, 1857–1862.

———. 1987. "How Einstein Discovered General Relativity: A Historical Tale With Some Contemporary Morals." In Malcolm A. H. MacCallum (ed.), *General Relativity and Gravitation: Proceedings of the 11th International Conference on General Relativity and Gravitation*. Cambridge: Cambridge University Press, 200–208.

———. 1989. "Einstein's Search for General Covariance, 1912–1915." In (Howard and Stachel 1989, 63–100).

———. 1993. "The Meaning of General Covariance: The Hole Story." In John Earman et al. (eds.), *Philosophical Problems of the Internal and External World, Essays on the Philosophy of Adolf Grünbaum*. (Konstanz: Universitätsverlag, Pittsburgh University Press, 129–160.

———. 1994a. "Scientific Discoveries as Historical Artifacts." In Kostas Gavroglu (ed.), *Current Trends in the Historiography of Science*. Dordrecht: Reidel, 139–148.

———. 1994b. "Changes in the Concepts of Space and Time Brought About by Relativity." In Carol C. Gould and Robert S. Cohen (eds.), *Artefacts, Representations and Social Practice*. Dordrecht: Kluwer, 141–162.

———. 1995. "History of Relativity." In Laurie M. Brown, Abraham Pais and Brian Pippard (eds.), *Twentieth Century Physics*, Vol. I. Bristol and Phila.: Institute of Physics Pub., New York: American Institute of Physics Press.

——— (ed.). 1998. *Einstein's Miraculous Year: Five Papers That Changed the Face of Physics*. Princeton: Princeton University Press.

———. 2003a. "Critical Realism: Bhaskar and Wartofsky." In Carol C. Gould (ed.), *Constructivism and Practice: Towards a Historical Epistemology*. Lanham, Md.: Rowman and Littlefield, 137–150.

———. 2003b. "Einstein's Intuition and the Post-Newtonian Approximation." Talk at the Mexico City Conference in Honor of Jerzy Plebanski, May 26, 2003. To appear in the Proceedings of the Conference.

———. 2005. "Fibered Manifolds, Geometric Objects, Structured Sets, G-Spaces and All That: The Hole Story from Space-Time to Elementary Particles." To appear in (Stachel forthcoming).

———. Forthcoming. *Going Critical: Selected Essays*. Dordrecht: Kluwer.

Struik, Dirk. 1933. "Outline of the History of Differential Geometry." *Isis* 19: 92–120, *Isis* 20: 161–191.

Weyl, Hermann. 1918a. "Reine Infinitesimalgeometrie." *Mathematische Zeitschrift* 2: 384–411. (English translation of excerpt given in this volume.)

———. 1918b. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. Berlin: Springer.

———. 1919. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 3rd ed.

———. 1921. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 4th ed.

———. 1923. *Raum-Zeit-Materie. Vorlesungen über die allgemeine Relativitätstheorie*. 5th ed.

Zuckerman, Harriet and Joshua Lederberg. 1986. "Forty years of genetic recombination in bacteria: postmature scientific discovery?" *Nature* 324: 629–631.

HERMANN GRASSMANN

ON THE RELATION OF NON-EUCLIDEAN GEOMETRY TO EXTENSION THEORY

Originally published as Appendix 1 (1877) to "A New Branch of Mathematics: The 'Ausdehnungslehre' of 1844 and Other Works" (Chicago: Open Court, 1995), pp. 279–280.

(Cf. §§15–23)[1]

To the detriment of science, the entire presentation in §§15–23 still remains almost totally unnoticed. Neither Riemann in his *Habilitationsschrift*[1] of 1854, first published in 1867, nor Helmholtz,[2] in his paper "Über die Tatsachen, welche der Geometrie zur Grunde liegen" (1868), nor even in his excellent lecture "Über den Ursprung und die Bedeutung der geometrischen Axiome" (1876) mention it, even though the foundations of geometry come into view much more simply than in those later publications.

In extension theory the straight line is quite special and, in contrast to Euclid, is the foundation for geometric definitions. In §16 the plane is defined as a collection of parallels that intersect a straight line, and space as a collection of parallels that intersect a plane; geometry can proceed no further, but the abstract science is not so limited. Since all points of a straight line may be numerically derived from two of its points, the straight line appears as a simple elementary domain of second order, and correspondingly the plane as a simple elementary domain of third, and ultimately space as one of fourth order.[3]

Thus for example the points of a plane are numerically derivable from three noncollinear points, e.g. by numbers x_1, x_2, x_3. Upon establishing a homogeneous equa-

1 Here is meant his *Habilitationsrede* "Über die Hypothesen, welche der Geometrie zu Grunde liegen," *Ges. Werke* 1st ed. p. 54ff, 2nd ed. p. 272ff.

2 The article appears in *Gott. Nachr.*, 1878, pp. 193–221, cf. also *Ges. wiss. Abh.*, vol. II., pp. 618–639. The lecture is found in his *Vortragen und Reden*, vol. II., p. 1 ff; Braunschweig: 1884.

3 To forestall confusion, I observe that the *displacements* in a plane form an elementary *extensive* domain of second order, those in space an elementary *extensive* domain of third order, and in general the displacements in a simple *elementary* domain of $(n + 1)$ -th order an elementary *extensive* domain of n-th order.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

tion between these three numbers, the collection of points satisfying this equation is reduced to a domain of second order. If this homogeneous equation is of first degree, then the domain so defined is elementary, that is a straight line; if however that equation is of higher degree it forms a curved line for which only some of the longimetric laws for the straight line are valid.

Turning to space, each of its points is numerically derivable from four points forming a tetrahedron, by four numbers $x_1, \ldots, x_4$. If, among these magnitudes, there exist two mutually independent homogeneous equations, neither of them of first degree, we then obtain doubly curved lines for which again only a part of the longimetric laws are true.

Now if we proceed another step beyond space, as a domain of fourth order, to a domain of fifth order (which does not exist geometrically), then one has five basis numbers $x_1, \ldots, x_5$, and if a homogeneous equation of first degree holds between them, then one returns to the simple elementary domain of fourth order, that is to Euclidean space. On the other hand, upon imposing on them a homogeneous equation of higher degree one also produces elementary domains of fourth order, but ones to which the Euclidean axioms no longer apply, and thus as it were to non-Euclidean spaces;[4] furthermore, one can proceed to an elementary domain of sixth order, and between the six determining numbers assume two higher homogeneous equations to obtain once more new elementary domains of fourth order,[5] and can in this way form an infinite sequence of non-Euclidean spaces, the equations of which immediately illuminate the extent to which the Euclidean axioms apply.

Thus extension theory also provides a fully adequate and completely general basis for these and similar considerations.

EDITORIAL NOTE

[1] The reference §§15–23 is to the body of Grassmann's text, which has not been reproduced in this book.

4 Thus for example resulting in Helmholtz's spherical space if one assumes a certain homogeneous equation of second degree between the five basis numbers mentioned above (or more generally a curved space upon adoption of an equation of arbitrary degree).

5 One could perhaps call such a space doubly curved, in contrast to the (simple) curved space just mentioned.

TULLIO LEVI-CIVITA

NOTION OF PARALLELISM ON A GENERAL MANIFOLD AND CONSEQUENT GEOMETRICAL SPECIFICATION OF THE RIEMANNIAN CURVATURE (EXCERPTS)

MEMORIA DI T. LEVI-CIVITA (PADOVA)

Originally published as "Nozione di parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura riemanniana" in Circolo Matematico di Palermo. Rendiconti, Vol. 42, 1916, pp. 173–204. Received 24 December 1916. Author's date: Padova, November 1916. The Introduction, §15, and the Critical Note are translated here.[1]

INTRODUCTION

Einstein's theory of relativity (now corroborated by the explanation of the famous secular inequality, revealed by observations on Mercury's perihelion, which was not predicted by Newton's law) considers the geometrical structure of space as very tenuously, but nonetheless intimately, dependent on the physical phenomena taking place in it, differently from classical theories, which assume the whole physical space as given *a priori*. The mathematical development of Einstein's grandiose conception (which finds in Ricci's absolute differential calculus its natural algorithmic instrument) utilizes as an essential element the curvature of a certain four-dimensional manifold and the Riemann symbols relative to it. Meeting these symbols—or, better said, continuously using them—in questions of such a general interest, led me to investigate whether it would be possible to somewhat reduce the formal apparatus commonly used in order to introduce them and to establish their covariant behaviour.[1]

Some progress in this direction is actually possible, and essentially forms the content of sections 15 and 16 of the present paper, which, initially conceived with this

1 Cf. e.g. L. Bianchi, *Lezioni di geometria differenziale*, Vol. I (Pisa, Spoerri, 1902), pp. 69–72.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

only purpose, gradually expanded to make some room for the geometric interpretation too.

At first, I thought I would undoubtedly find it [that interpretation] in Riemann's original works "*Über die Hypothesen welche [der] Geometrie zu Grunde liegen*" and "*Commentatio Mathematica…*,"[2] but only an embryo of it can be found there. Indeed, on the one hand, looking closely at the quoted sources, one gets the impression that Riemann had actually in mind the characterization of the intrinsic and
[174] invariant curvature that shall be specified here (sections 17–18). | On the other hand, neither in Riemann nor in Weber's explicative comment,[3] is to be found a trace of those specifications (notion of parallel directions on a general manifold and consideration of a geodetic infinitesimal quadrangle with two parallel sides), that we shall recognize as indispensable from the geometrical point of view. Moreover, one cannot—or at least, I was not able to—justify the formal step in terms of which, according to Riemann, from the premises, which are impeccable, one should obtain an equivalently impeccable final expression of the curvature.

I will present to the reader this doubt of mine, providing him with the elements required to form an opinion in a final critical note.

The first and more extended part of the paper (sections 1–14) is devoted to an introduction and an illustration of the notion of parallelism in a V_n with any metric.

One begins with the infinitesimal field, trying to characterize the parallelism of two directions $(\alpha), (\alpha')$ through two very close points P and P'. To this purpose, one should remember that any manifold V_n can be looked at as embedded in an Euclidean space S_N of a sufficiently high number N of dimensions, and notice, first of all, that, for any direction (f) of S_N through P, ordinary parallelism would require, in such a space,

$$\text{angle}(f)\widehat{(}\alpha) = \text{angle}(f)\widehat{(}\alpha')$$

for any (f). Now, parallelism in V_n is defined limiting oneself to require that the condition be satisfied *for all the* (f) *belonging to* V_n (namely to the set of directions of S_N tangent to V_n in P).

In order to justify this definition, it should be noted that, while for an Euclidean V_n, it reproduces, as is necessary, the elementary behaviour, it has in any case an intrinsic character, since it ultimately turns out to depend only on the metric of V_n, and not on the auxiliary ambient space S_N as well. Indeed, the analytic version of our definition of parallelism is realized as follows: Once V_n is given general coordinates x_i $(i = 1, 2, \ldots, n)$, , let dx_i be the increments corresponding to the displacement from P to P'; $\xi^{(i)}$ the parameters of a generic direction (α) through P; $\xi^{(i)} + d\xi^{(i)}$ those belonging to an infinitely close direction (α') through P'. The condition of parallelism is expressed by the n equations

2 B. Riemann, *Gesammelte mathematische Werke* (Leipzig, Teubner, 1876), pp. 261–263, 381–382.

3 loc. cit.[2] pp. 384–389.

$$d\xi^{(i)} + \sum_{j,l=1}^{n} \left\{ \begin{matrix} jl \\ i \end{matrix} \right\} dx_j \, \xi^{(l)} = 0 \qquad (i = 1, 2, \ldots, n), \tag{A}$$

where the $\left\{ \begin{matrix} jl \\ i \end{matrix} \right\}$ denote well-known Christoffel symbols.

Once the law by means of which one goes from one point to an infinitely close point is acquired, one is provided with all the means required in order to perform the transport of parallel directions along any curve C. If $x_i = x_i(s)$ are its parametric equations, | one only needs to consider in eqs. (A) the x_i and subordinately the [175]

$$\left\{ \begin{matrix} jl \\ i \end{matrix} \right\},$$

as assigned functions, the $\xi^{(i)}$ as functions to be determined of the parameter s, and one has the ordinary linear system

$$\frac{d\xi^{(i)}}{ds} + \sum_{j,l=1}^{n} \left\{ \begin{matrix} jl \\ i \end{matrix} \right\} \frac{dx_j}{ds} \, \xi^{(l)} = 0 \qquad (i = 1, 2, \ldots, n)$$

reducible to a typical form (said "with hunched determinant"), which already appeared in other researches and was the object of a systematic investigation by Mr. Eiesland,[4] Laura,[5] Darboux,[6] Vessiot.[7]

Here is some geometrical consequence.

1. The direction through a generic point P parallel to a direction (α) through any other point P_0 depends in general on the path followed from P_0 to P. Independence from the path is an exclusive property of Euclidean manifolds.

2. Along a given geodesic, directions of the tangents are parallel, a result that generalizes an obvious feature of the straight line in a Euclidean space (the one that Euclid himself sets as a primordial intuitive notion of a straight line at the beginning of *Elements*).

3. The parallel transport along any path of two concurrent directions preserves their angle. By this we obviously mean that the angle formed by two generic direc-

---

4 J. Eiesland, "On the Integration of a System of Differential Equations in Kinematics" *American Journal of Mathematics*, vol. XXVIII (1906), pp. 17–42.

5 E. Laura, "Sulla integrazione di un sistema di quattro equazioni differenziali lineari a determinante gobbo per mezzo di due equazioni di Riggati" *Atti della Accademia delle Scienze di Torino*, vol. XLII, 1906–1907, pp. 1089–1108; vol. XLII, 1907–1908, pp. 358–378.

6 G. Darboux, "Sur certains systèmes d'equations linéaires*" Comptes rendu hebdomadaires des séances de l'Académie des Sciences*, t. CXLVIII (1er semestre 1909), pp. 332–335, and "Sur les systèmes d' équations différentielles homogènes" (*Ibid*., pp. 673–679 and pp. 745–754.

7 E. Vessiot, "Sur l'intégration des systèmes linéaires à déterminant gauche" *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, t. CXLVIII (1er semestre 1909), pp. 332–335.

tions through the same point is also the angle formed by their parallels through another point. Taking into account the mentioned property of the geodesics, one derives as a corollary that, along a geodesic, parallel directions are always equally inclined with respect to the geodesic itself. If in particular one deals with a V_2, this condition is also sufficient; hence, for ordinary surfaces, parallelism along a geodesic is equivalent to isogonality.

I am not specifying how the content is arranged in the various sections. A look at the summary at the end of the paper will supply the necessary information.

[…]

[195]

§15.
2° ORDER DIFFERENTIALS - INVARIANT DETERMINATIONS - RICCI'S LEMMA.

In a given investigation, let the independent variables, for instance $x_1, x_2, \ldots, x_n$, be fixed. As is known from the calculus, it is always legitimate to consider the second order differentials $d^2x_1, d^2x_2, \ldots, d^2x_n$ as vanishing. Such a convention, however, does not have an invariant character with respect to changes of variables. Indeed, if the x_i are replaced by n independent combinations thereof $\chi_i(x_1, x_2, \ldots, x_n)$, the second differentials

$$d^2\chi_i = \sum_{j,l=1}^{n} \frac{\partial^2 \chi_i}{\partial x_j \, \partial x_l} dx_j \, dx_l$$

(computed on the basis of the hypothesis $d^2x_i = 0$) turn out to be in general different from zero.

If to the variables a quadratic differential form is associated, referring for instance to the metric of a V_n (in the notations of the preceding sections), the way to an invariant characterization is facilitated. It is sufficient to assume the d^2x_i (not vanishing, but) defined as follows:

$$d^2x_i + \sum_{j,l=1}^{n} \left\{ \begin{matrix} jl \\ i \end{matrix} \right\} dx_j \, dx_l = 0 \qquad (i = 1, 2, \ldots, n).$$

From the geodesic equations (sec. 7), multiplied by ds^2, it appears that such d^2x_i are those belonging to the variables along the geodesic through the generic point $(x_1, x_2, \ldots, x_n)$ in the similarly generic direction $(dx_1, dx_2, \ldots, dx_n)$. This geometric interpretation guarantees *a priori* that the above convention has the desired invariant character, making unnecessary a material check, which, on the other hand, could be done straightforwardly.

Similarly for the superposition of two independent systems of increments dx_i and δx_i, one might set $d\delta x_i = \delta dx_i = 0$, but, while the invertibility of the increments d

and δ has, as is easily checked, an invariant character, the same does not hold as far as setting $d\delta x_i = 0$ is concerned. We shall replace them by:

$$d\delta x_i + \sum_{j,l=1}^{n} \begin{Bmatrix} jl \\ i \end{Bmatrix} dx_i \delta x_j = 0, \tag{31}$$

which imply

$$d\,\delta x_i = \delta\, dx_i, \tag{31'}$$

and contain, as a particular case, for $d = \delta$, the previous expressions for the $d^2 x_i$.

The invariance of eqs. (31) with respect to changes of variables can be derived from the geometric interpretation as well. One only needs to observe that, writing $\delta x_i = \varepsilon \xi_i^{(i)}$ (with ε an infinitesimal constant), eqs. (31) become identical with eqs. (I_a), so that they express how the δx_i must be altered, as a consequence of the displacement $(dx_1, dx_2, \ldots, dx_n)$, | in order that they define directions parallel to one [196] another. This invariant property, besides verifying it directly, could be controlled with an elegant formal device sketched by Riemann[8] and made explicit by Weber.[9]

From eqs. (31), taking into account

$$d\,a_{ik} = \sum_{j=1}^{n} \frac{\partial a_{ik}}{\partial x_j} dx_j = \frac{1}{2} \sum_{j=1}^{n} (a_{ij,k} + a_{jk,i}) dx_j = \frac{1}{2} \sum_{j,l=1}^{n} \left[a_{lk} \begin{Bmatrix} ij \\ l \end{Bmatrix} + a_{li} \begin{Bmatrix} jk \\ l \end{Bmatrix} \right] dx_j,$$

it follows identically[2]

$$d \sum_{i,k=1}^{n} a_{ik}\, \delta x_h\, \delta x_k = 0, \tag{32}$$

as well as

$$d \sum_{i,k=1}^{n} a_{ik}\, dx_h\, \delta x_k = 0.$$

These relations are equivalent to the well-known result of the absolute differential calculus that the covariantly derived system of the coefficients a_{ik} of the fundamental form vanish identically (Ricci's lemma).

[...]

1. CRITICAL NOTE [201]

We have already pointed out in sec. 15, the expressions (31)

8 loc. cit.[2] p. 381.
9 loc. cit.[2] p. 388.

$$d\delta x_i + \sum_{j,l=1}^{n} \left\{ \begin{matrix} jl \\ i \end{matrix} \right\} dx_j\, \delta x_l = 0$$

[202] | of the second order differentials do not differ from those which are arrived at by making Riemann's comprehensive definition explicit.

From the same section it is also deduced that, with these expressions of the second differentials, one has identically (Ricci's lemma)

$$\delta\, ds^2 = d\,\delta s^2 = d\Phi = \delta\Phi = 0, \tag{48}$$

where

$$\Phi = \sum_{i,k=1}^{n} a_{ik}\, dx_i \delta x_k,$$

and ds^2, δs^2 stand, of course, respectively for

$$\sum_{i,k=1}^{n} a_{ik}\, dx_i\, dx_k \qquad \text{and} \qquad \sum_{i,k=1}^{n} a_{ik}\, \delta x_i\, \delta x_k.$$

In this context the meaning to be attributed to the trinomial considered by Riemann:

$$R = \delta^2 ds^2 - 2\, d\delta\Phi + d^2\, \delta s^2$$

seems unambiguous, and such meaning, by virtue of (48), implies necessarily $R = 0$.

Riemann states[10] instead that: "Haec expressio (that is R) invenietur $= J$" (J having the value (45)). Weber, in his elucidations, dwells on the way the second differentials are introduced,[11] but, after deriving their explicit expression, simply says:[12] "woraus man leicht den Ausdruck erhält $R = J$."

Probably, there is just some blemish in Riemann's explicit expression for R that obscures the concept. I flatter myself that I have substantially reconstructed such a concept, but I was not able to adjust the symbol. If this can be achieved, it will be the case to pay full tribute, on this point too, to Riemann's genius.

---

10 loc. cit.[2] p. 381

11 Adding, with no further justification, the supplementary conditions

$$d^2\delta s^2 = \delta^2 ds^2 = -2 d\delta\Phi.$$

By virtue of (48) (and provided the formulae are read as they are actually written) everything vanishes.

12 loc. cit.[2] p. 388

I shall end with an observation about the calculation of the curvature with reference to particular variables, which is indicated by Riemann[13] and developed by Weber.[14] Here is, to begin with, what the matter is about.

Let us choose coordinates $x_1, x_2, \ldots, x_n$ such that, *at a given point* P, all the symbols

$$\left\{ \begin{matrix} jl \\ i \end{matrix} \right\}$$

vanish (which is always possible, as was pointed out by Weber). Let us consider two independent sets of differentials $dx_i, \delta x_i$, considering | all the second differentials [203]
$d^2x_i, d\delta x_i, \delta dx_i, \delta^2 x_i$ as vanishing. Let P' and Q denote the points of coordinates $x_i + dx_i, x_i + \delta x_i$, and a'_{hk} the coefficients of the squared line element in P'. Set, in particular,

$$(\delta s^2)_{P'} = \sum_{h,k=1}^{n} a'_{hk}\, \delta x_h\, \delta x_k,$$

let us apply to the a' the Taylor expansion with respect to the increments d up to the second order. In such approximation one has

$$(\delta s^2)_{P'} = \delta s^2 + \frac{1}{2} \sum_{h,k,j,l=1}^{n} \frac{\partial^2 a_{hk}}{\partial x_j\, \partial x_l}\, dx_j\, dx_l\, \delta x_h\, \delta x_k,$$

δs^2 and the second derivatives referring, of course, to P. As shown by Weber, due to the way the variables were fixed, special relations hold between the values of the second derivatives of the a_{hk} in P. Taking them into account, one finds, with some manipulation,

$$(\delta s^2)_{P'} = \delta s^2 + \frac{1}{3} \sum_{h,k,j,l=1}^{n} \left[\frac{\partial^2 a_{hk}}{\partial x_j \partial x_l} + \frac{\partial^2 a_{jl}}{\partial x_h \partial x_k} - \frac{\partial a_{hj}}{\partial x_k \partial x_l} - \frac{\partial a_{kl}}{\partial x_h \partial x_j} \right] dx_j\, dx_l \delta x_h \delta x_k.$$

The sum can be looked at as the expression which, as the basis of formula (45), is taken on by $-I$, when variables x specified as above are adopted.

Therefore, taking into account (47), we derive

$$\frac{(\delta s^2)_{P'} - \delta s^2}{(ds\, \delta s \sin\psi)^2} = -\frac{1}{3} K, \tag{49}$$

which Riemann, in the quoted passage, states in words (multiplying both sides by 4, in order to show up the area of the triangle $PP'Q$ in the denominator).

---

13 loc. cit.[2] p. 261

14 loc. cit.[2] pp. 384–387

I come, at last, to my point:

If Q^* denotes the extremum of the line element $(\delta s^2)_{P'}$ (corresponding to the increments δx_i), eq. (49) can be written

$$\frac{\overline{P'Q^*}^2 - \overline{PQ}^2}{(ds\delta s\sin\psi)^2} = -\frac{1}{3}K; \tag{49'}$$

whereas eq. (46) (with an overall change of sign) reads

$$\frac{\overline{P'Q'}^2 - \overline{PQ}^2}{(ds\delta s\sin\psi)^2} = -K. \tag{46'}$$

As can be seen, the right-hand sides are in the ratio 1 to 3. The lack of coincidence is manifestly due to the fact that the point Q' (fourth vertex of the parallelogrammoid), which is reached through the invariant procedure, is well distinct from Riemann's point Q^*, analytically defined with reference to particular variables.

To localize the discrepancy about the formulae, it helps to work out our procedure too (as is of course allowed given its invariant character) in Riemann's special vari-
[204] ables. | Eqs. (31) give then, in so far as they refer to the point P,

$$d^2x_i = d\delta x_i = \delta dx_i = \delta^2 x_i = 0;$$

but it does not follow that the higher differentials, such as δd^2x_i, $d^2\delta x_i$, etc. must vanish at the same point as well. Riemann's calculation on the contrary is based on the hypothesis that all differentials of an order higher than the first must vanish: a legitimate hypothesis too, but not one endowed with an invariant character (with respect to changes of variables). Therefore, it should not come as a surprise that the results are different: one should rather notice the fortuitous analogy between formulae $(49')$ and $(46')$, whose right-hand sides differ only by a numerical factor.

EDITORIAL NOTES

[1] This text has been translated by Silvio Bergia.

[2] In the original text, the index $_h$ is missing from the expression δx_h in eq. (32).

HERMANN WEYL

PURELY INFINITESIMAL GEOMETRY (EXCERPT)

Originally published as "Reine Infinitesimalgeometrie" in Mathematische Zeitschrift 2, 1918, pp. 384–411. Excerpt covers pp. 384–401.

1. INTRODUCTION: CONCERNING THE RELATION BETWEEN GEOMETRY AND PHYSICS

The real world, into which we have been placed by virtue of our consciousness, *is not there* simply and all at once, but *is happening*; it passes, annihilated and newly born at each instant, a continuous one-dimensional succession of states in *time*. The arena of this temporal happening is a three-dimensional Euclidean *space*. Its properties are investigated by *geometry*, the task of *physics* by comparison is to conceptually comprehend the real that exists in space and to fathom the laws persisting in its fleeting appearances. Therefore, physics is a science which has geometry as its foundation; the concepts however, through which it represents reality—matter, electricity, force, energy, electromagnetic field, gravitational field, etc.—belong to an entirely different sphere than the geometrical.

This old view concerning the relation between the form and the content of reality, between geometry and physics, has been overturned by Einstein's theory of relativity.[1] The *special theory of relativity* led to the insight that space and time are fused into an indissoluble whole which shall here be called the *world*; the world, according to this theory, is a four-dimensional Euclidean manifold—Euclidean with the modification that the underlying quadratic form of the world metric is not positive definite but is of inertial index 1. The *general theory of relativity*, in accordance with the spirit of modern physics of local action [*Nahewirkungsphysik*], admits that as valid only in the infinitely small, hence for the world metric it makes use of the more general concept of a metric [*Maßbestimmung*] based on a quadratic *differential* form, developed by Riemann in his habilitation lecture. | But what is new in principle in this [385] is the insight that the metric is not a property of the world in itself, rather, spacetime as the form of appearances is a completely formless four-dimensional continuum in

1 I refer to the presentation in my book *Raum, Zeit, Materie*, Springer 1918 (in the sequel cited as RZM), and the literature cited there.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

the sense of analysis situs. The metric, however, expresses something real that exists in the world, which produces physical effects on matter by means of centrifugal and gravitational forces, and whose state is in turn determined according to natural laws by the distribution and composition of matter. By removing from Riemannian geometry, which claims to be a purely "local geometry," [*Nahe-Geometrie*] an inconsequence still currently adhering to it, ejecting one last element of non-local geometry [*ferngeometrisches Element*] which it had carried along from its Euclidean past, I arrived at a world metric from which not only arises gravitation, but also the electromagnetic effects, and therefore, as one may assume with good reason, accounts for all physical processes.[2] According to this theory, *everything real that exists in the world is a manifestation of the world metric*; the physical concepts are none other than the geometric ones. The only difference that exists between geometry and physics is that geometry fathoms in general what lies in the nature of the metric concepts,[3] whereas physics has to determine the law by which the real world is distinguished among all the four-dimensional metric spaces possible according to geometry and pursue its consequences.[4]

In this note, I want to develop that *purely infinitesimal geometry* which, according to my conviction, contains the physical world as a special case. The construction of the local geometry proceeds adequately in three steps. On the first step stands the *continuum* in the sense of analysis situs, without any metric—physically speaking, *the empty world*; on the second the *affinely connected continuum*—I so call a manifold in which the concept of infinitesimal parallel displacement of vectors is meaningful; in [386] | physics, the affine connection appears as *the gravitational field*—; finally on the third, the *metric* continuum—physically: *the "aether,"* whose states are manifested in the phenomena of matter and electricity.

2. SITUS-MANIFOLD (EMPTY WORLD)

As a consequence of the difficulty in grasping the intuitive character of the continuous connection by means of a purely logical construction, a completely satisfactory analysis of the concept of an *n-dimensional manifold* is not possible today.[5] The following is sufficient for us: An n-dimensional manifold refers to n coordinates $x_1\ x_2\ \ldots\ x_n$, of which each possesses at each point of the manifold a particular numerical value: different sets of values of the coordinates correspond to different

2 A first communication about this appeared under the title "Gravitation und Elektrizität" in *Sitzungsber. d. K. Preuß. Akad. d. Wissenschaften* 1918, p. 465.

3 Naturally, traditional geometry leaves the path of this, its principal task, and immediately takes on the less specific one by not making space itself anymore the object of its investigation, but the structures possible in space, special classes and their properties they are endowed with on the basis of the space-metric.

4 I am bold enough to believe that the totality of physical phenomena can be derived from a single universal world law of greatest mathematical simplicity.

5 See also H. Weyl, *Das Kontinuum* (Leipzig 1918), specifically pp. 77 ff.

points; if $\bar{x}_1\ \bar{x}_2\ \ldots\ \bar{x}_n$ is a second system of coordinates then there exist between the x- and the $\bar{x}$-coordinates of the same arbitrary point regular relations

$$x_i = f_i(\bar{x}_1\ \bar{x}_2\ \ldots\ \bar{x}_n) \qquad (i = 1, 2, \ldots, n),$$

where f_i denote purely logically-arithmetically constructible functions; of these we presuppose not only that they are continuous, but also that they posses continuous derivatives

$$\alpha_{ik} = \frac{\partial f_i}{\partial \bar{x}_k},$$

whose determinant does not vanish. The last condition is necessary and sufficient for the affine geometry to be valid in the infinitely small, namely that there exist invertible linear relationships between the coordinate differentials in the two systems:

$$dx_i = \sum_k \alpha_{ik} d\bar{x}_k. \tag{1}$$

We assume the existence and continuity of higher order differentials where required during the course of the investigation. In any case, the concept of the continuous and continuously differentiable point-function, if necessary also the $2, 3, \ldots$ times continuously differentiable, has therefore an invariant meaning independent of the coordinate system. The coordinates themselves are such functions. An n-dimensional manifold for which we regard no properties other than those lying within the concept of an n-dimensional manifold, we call—in physical terminology—an (n-dimensional) *empty world*. |

The relative coordinates dx_i of a point $P' = (x_i + dx_i)$ infinitely close to the point [387]
$P = (x_i)$ are the components of a *line element* in P, or an *infinitesimal displacement* $\overrightarrow{PP'}$ of P. In going to a different coordinate system the formulae (1) apply for these components, the α_{ik} denoting the corresponding derivatives at the point P. More generally, on the basis of a definite coordinate system in the neighborhood of P, any n numbers ξ^i $(i = 1, 2, \ldots, n)$ given in a definite order, characterize at the point P a *vector* (or a *displacement*) at P. The components ξ^i respectively $\bar{\xi}^i$ of the same vector in any two coordinate systems, the "unbarred" one and the "barred" one, are related by the same linear transformation equations (1):

$$\xi^i = \sum_k \alpha_{ik} \bar{\xi}^k.$$

Vectors at P can be added and multiplied by numbers; thus they form a "linear" or "affine" totality [*Gesamtheit*]. With each coordinate system are associated n "unit vectors" e_i at P, namely those vectors which in the coordinate system in question have the components

e_1	1, 0, 0, ..., 0
e_2	0, 1, 0, ..., 0
..	... ...
e_n	0, 0, 0, ..., 1

Any two (linearly independent) line elements at P with the components dx_i and δx_i respectively span a (two-dimensional) area element at P with the components

$$dx_i \delta x_k - dx_k \delta x_i = \Delta x_{ik},$$

each three (independent) line elements dx_i, δx_i, $\mathfrak{d} x_i$ at P, a (three-dimensional) volume element with the components

$$\begin{vmatrix} dx_i & dx_k & dx_l \\ \delta x_i & \delta x_k & \delta x_l \\ \mathfrak{d} x_i & \mathfrak{d} x_k & \mathfrak{d} x_l \end{vmatrix} = \Delta x_{ikl};$$

etc. A linear form depending on an arbitrary line- or area- or volume- or ... element at P is called a *linear tensor* of order 1, 2, 3… respectively. By using a particular coordinate system, the coefficients a of this linear form

$$\sum_i a_i dx_i, \quad \text{resp.} \quad \frac{1}{2!}\sum_{ik} a_{ik}\Delta x_{ik}, \quad \frac{1}{3!}\sum_{ikl} a_{ikl}\Delta x_{ikl}, \ ...$$

[388] | can be uniquely normalized through the alternation requirement; e.g., for the case just written down this implies that the triple of indices (ikl), which arise through an even permutation of itself corresponds to the same coefficient a_{ikl}, whereas under odd permutations the coefficient changes into its negative, that is

$$a_{ikl} = a_{kli} = a_{lik} = -a_{kil} = -a_{lki} = -a_{ilk}.$$

The coefficients normalized in this manner are called the *components* of the tensor in question. From a scalar field f one obtains through differentiation a linear tensor field of order 1 with the components

$$f_i = \frac{\partial f}{\partial x_i};$$

from a linear tensor field f_i of order 1, one of 2nd order:

$$f_{ik} = \frac{\partial f_i}{\partial x_k} - \frac{\partial f_k}{\partial x_i};$$

from one of order 2, a linear tensor field of order 3:

$$f_{ikl} = \frac{\partial f_{kl}}{\partial x_i} + \frac{\partial f_{li}}{\partial x_k} + \frac{\partial f_{ik}}{\partial x_l};$$

etc. These operations are independent of the coordinate system used.[6]

A linear tensor of the 1st order at P we will call a *force* acting there. Assuming a definite coordinate system, such a force is thus characterized by n numbers ξ_i, which transform contragrediently to the components of the displacement under a change to another coordinate system:

$$\xi_i = \sum_k \alpha_{ki} \xi_k.$$

If η^i are the components of an arbitrary displacement at P, then

$$\sum_i \xi_i \eta^i$$

is an invariant. By a *tensor* at P, one generally understands a linear form of one or more arbitrary displacements and forces at P. For example, if we are dealing with a linear form of three arbitrary displacements ξ, η, ζ and two arbitrary forces ρ, σ:

$$\sum a_{ikl}^{pq} \xi^i \eta^k \zeta^l \rho_p \sigma_q,$$

then we speak of a tensor of order 5, with the components a being covariant with respect to the indices ikl and contravariant with respect to the indices pq. A displacement is itself a contravariant | tensor of 1st order, the force a covariant one. The [389]
fundamental operations of tensor algebra are:[7]

1. Addition of tensors and multiplication by a number;
2. Multiplication of tensors;
3. Contraction.

Accordingly, tensor algebra can already be constructed in the empty world—it does not presuppose any metric [*Maßbestimmung*]—of tensor analysis, however, only that of "linear" tensors.

A "*motion*" in our manifold is given, if to each value s of a real parameter is assigned a point in a continuous manner; by using the coordinate system x_i, the motion is expressed by the formulae $x_i = x_i(s)$, in which the x_i on the right are to be understood as function symbols. If we presuppose continuous differentiability, then we obtain, independently of the coordinate system, for each point $P = (s)$ of the motion a vector at P with the components:

6 RZM, §13.
7 RZM, §6.

$$u^i = \frac{dx_i}{ds},$$

the *velocity*. Two motions, arising from one another through continuous monotonic transformation of the parameter s describe the same *curve*.

3. AFFINELY CONNECTED MANIFOLD (WORLD WITH GRAVITATIONAL FIELD)

3.1 The Concept of the Affine Connection

If P' is infinitely close to the fixed point P, then P' is *affinely connected with* P, if for each vector at P it is determined into which vector at P' it will transform under *parallel displacement* from P to P'. The parallel displacement of all vectors at P from there to P' must evidently satisfy the following requirement.

A. *The transfer of the totality of vectors from* P *to the infinitely close point* P' *by means of parallel displacement produces an affine transformation of the vectors at* P *to the vectors at* P'.

If we use a coordinate system in which P has the coordinates x_i, P' the coordinates $x_i + dx_i$, an arbitrary vector at P the components ξ^i, and the vector at P', that results from it through parallel displacement to P', the components $\xi^i + d\xi^i$, then
[390] $d\xi^i$ must therefore depend linearly on the ξ^i: |

$$d\xi^i = -\sum_r d\gamma^i{}_r \xi^r.$$

$d\gamma^i{}_r$ are infinitesimal quantities which depend only on the point P and the displacement $\overrightarrow{PP'}$ with the components dx_i, but not on the vector ξ subject to parallel displacement. From now on, we consider affinely connected manifolds; in such a manifold, each point P is affinely connected to all its infinitely close points. A second requirement is still to be imposed on the concept of parallel displacement, that of *commutativity*.

B. *If* P_1, P_2 *are two points infinitely close to* P *and if the infinitesimal vector* $\overrightarrow{PP_1}$ *becomes* $\overrightarrow{P_2P_{21}}$ *under parallel displacement from* P *to* P_2, *and* $\overrightarrow{PP_2}$ *becomes* $\overrightarrow{P_1P_{12}}$ *under parallel displacement to* P_1, *then the points* P_{12} *and* P_{21} *coincide*. (An infinitely small parallelogram results.)

If we denote the components of $\overrightarrow{PP_1}$ by dx_i, and those of $\overrightarrow{PP_2}$ by δx_i, then the requirement in question obviously implies that

$$d\delta x_i = -\sum_r d\gamma^i{}_r \cdot \delta x_r \tag{2}$$

is a symmetric function of the two line elements d and δ. Consequently, $d\gamma^i{}_r$ must be a linear form of the differentials dx_i,

$$d\gamma^i{}_r = -\sum_s \Gamma^i{}_{rs}\, dx_s,$$

and the coefficients Γ, the "*components of the affine connection*," which depend only on the location of P, must satisfy the symmetry condition

$$\Gamma^i{}_{sr} = \Gamma^i{}_{rs}.$$

Because of the way in which the infinitesimal quantities are dealt with in the formulation of the requirement B, it could be objected that the latter lacks a precise meaning. Therefore, we want to determine explicitly through a rigorous proof that the symmetry of (2) is a condition independent of the coordinate system. For this purpose, we make use of a (twice differentiable) scalar field f. From the formula for the total differential

$$df = \sum_i \frac{\partial f}{\partial x_i} dx_i$$

we infer, that if ξ^i are the components of an arbitrary vector at P, |

$$df = \sum_i \frac{\partial f}{\partial x_i} \xi^i \qquad [391]$$

is an invariant independent of the coordinate system. We form its variation under a second infinitesimal displacement δ, in which the vector ξ shall be displaced parallel to itself from P to P_2, and obtain

$$\delta df = \sum_{ik} \frac{\partial^2 f}{\partial x_i \partial x_k} \xi^i \delta x_k - \sum_{ir} \frac{\partial f}{\partial x_i} \cdot d\gamma^i{}_r \xi^r.$$

If we replace in this expression ξ^i again by dx_i and subtract from this equation the one obtained by interchanging d and δ, then the invariant

$$\Delta f = (\delta d - d\delta) f = \sum_i \left\{ \frac{\partial f}{\partial x_i} \sum_r (d\gamma^i{}_r \delta x_r - \delta\gamma^i{}_r dx_r) \right\}.$$

results. The relations

$$\sum_r (d\gamma^i{}_r \delta x_r - \delta\gamma^i{}_r dx_r) = 0$$

contain the necessary and sufficient condition that for any scalar field f the equation $\Delta f = 0$ is satisfied.

In physical terms, an affinely connected continuum is to be described as a world in which a *gravitational field* exists. The quantities $\Gamma^i{}_{rs}$ are the components of the gravitational field. The formulae, according to which these components transform in changing from one coordinate system to another, we need not state here. Under linear transformations the $\Gamma^i{}_{rs}$ behave with respect to r and s like the covariant components of a tensor and with respect to i like the contravariant components, but lose this character under non-linear transformations. However, the changes $\delta\Gamma^i{}_{rs}$, which are experienced by the quantities Γ, if one arbitrarily varies the affine connection of the manifold, form the components of a generally-invariant tensor of the given character.

What is to be understood by *parallel displacement of a force* at P from there to the infinitely close point P' results from the requirement that the invariant product of this force and an arbitrary vector at P is preserved under parallel displacement. If ξ_i are the components of the force, η^i those of the displacement, then[8]

$$d(\xi_i\eta^i) = (d\xi_i \cdot \eta^i) + \xi_r d\eta^r = (d\xi_i - d\gamma^r{}_i\xi_r)\eta^i = 0$$

yields the formula

$$d\xi_i = \sum_r d\gamma^r{}_i\xi_r.$$

[392] | At each point P, one can introduce a coordinate system x_i of a kind—I call it *geodesic* at P—such that in it, the components of the affine connection $\Gamma^i{}_{rs}$ vanish at the point P. If x_i are initially arbitrary coordinates that vanish at P, and $\Gamma^i{}_{rs}$ designate the components of the affine connection at the point P in this coordinate system, then one obtains a geodesic coordinate system $\bar{x}_i$ via the transformation

$$x_i = \bar{x}_i - \frac{1}{2}\sum_{rs}\Gamma^i{}_{rs}\bar{x}_r\bar{x}_s. \tag{3}$$

Namely, if we consider the $\bar{x}_i$ as independent variables and their differentials $d\bar{x}_i$ as constants, then one has in the sense of Cauchy at $P(\bar{x}_i = 0)$:

$$dx_i = d\bar{x}_i, \qquad d^2x_i = -\Gamma^i{}_{rs}d\bar{x}_r d\bar{x}_s,$$

therefore,

$$d^2x_i + \Gamma^i{}_{rs}d\bar{x}_r d\bar{x}_s = 0.$$

Because of their invariant nature, the last equations in the coordinate system $\bar{x}_i$ become:

$$d^2\bar{x}_i + \bar{\Gamma}^i{}_{rs}d\bar{x}_r d\bar{x}_s = 0.$$

---

8 In the following we will use Einstein's convention that summation is always to be carried out over indices which occur twice in a formula without our finding it necessary to always place a summation sign in front of it.

For arbitrary constant $d\bar{x}_i$ these are, however, satisfied only if all the $\bar{\Gamma}^i{}_{rs}$ vanish. *Therefore, through an appropriate choice of the coordinate system, the gravitational field can always be made to vanish at a single point*. Through the requirement of "geodesy" at P the coordinates in the neighborhood of P are determined up to linear transformation excluding terms of third order; i.e., if x_i, $\bar{x}_i$ are two coordinate systems geodesic at P, and if the x_i as well as the $\bar{x}_i$ vanish at P, then by neglecting terms in $\bar{x}_i$ of order 3 and higher, linear transformation equations $x_i = \sum_k \alpha_{ik}\bar{x}_k$ with constant coefficients α_{ik} apply.

3.2 Tensor Analysis, Straight Line

Only in an affinely connected space can *tensor analysis* be fully established. If for example f_i^k are the components of a 2nd order tensor field, covariant in i and contravariant in k, then with the aid of an arbitrary displacement ξ and a force η at the point P, we form the invariant

$$f_i^k \xi^i \eta_k$$

and its change under an infinitely small displacement d of the point P, in which ξ and η are displaced parallel with respect to themselves. We have

$$d(f_i^k \xi^i \eta_k) = \frac{\partial f_i^k}{\partial x_l} \xi^i \eta_k dx_l - f_r^k \eta_k d\gamma^r{}_i \xi^i + f_i^r \xi^i d\gamma^k{}_r \eta_k,$$

| and therefore [393]

$$f_{il}^k = \frac{\partial f_i^k}{\partial x_l} - \Gamma_{il}^r f_r^k + \Gamma_{rl}^k f_i^r$$

are the components of 3rd order tensor field, covariant in il and contravariant in k, which arises from the given 2nd order tensor field in a coordinate independent manner.

In the affinely connected space, the concept of *straight or geodesic line* gains a definite meaning. The straight line arises as the trajectory of the initial point of the vector which is displaced in its own direction keeping it parallel to itself; it can therefore be described as that curve the direction of which remains unchanged. If u^i are the components of that vector, then during the course of the motion the equations

$$du^i + \Gamma^i{}_{\alpha\beta} u^\alpha dx_\beta = 0,$$
$$dx_1 : dx_2 : \ldots : dx_n = u^1 : u^2 : \ldots : u^n$$

should always hold. The parameter s used in describing the curve can thus be normalized in such a way that

$$\frac{dx_i}{ds} = u^i$$

identically along s, and the differential equations of the straight line are then

$$w^i \equiv \frac{d^2x_i}{ds^2} + \Gamma^i_{\alpha\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0.$$

For each arbitrary motion $x_i = x_i(s)$, the left hand sides of these equations are the components of a vector invariantly linked to the motion at the point s, the acceleration. Actually, if ξ_i is an arbitrary force at that point, which during the transition to the point $s + ds$ is displaced parallel to itself, then

$$\frac{d(u^i\xi_i)}{ds} = w^i\xi_i.$$

A motion, whose acceleration vanishes identically is called a *translation*. A straight line—this is another way of grasping our above explanation—is to be understood as the trajectory of a translation.

3.3 Curvature

If P and Q are two points connected by a curve, and a vector is given at the first point, then one can displace this vector parallel to itself along the curve from P to Q. The resulting *vector transfer* is however in general *not integrable*; i.e. the vector |
[394] which one ends up with at Q depends on the path along which the transport takes place. Only in the special case of integrability does it make sense to speak of the *same* vector at two different points P and Q; these are understood to be vectors which arise from one another under parallel transport. In this case, the manifold is called *Euclidean*. In such a manifold, special "linear" coordinate systems can be introduced which are distinguished by the fact that equal vectors at different points have equal components. Any two such linear coordinate systems are related by linear transformation equations. In a linear coordinate system the components of the gravitational field vanish identically.

On the infinitely small parallelogram constructed above (§3, I., B.), we attach at the point P an arbitrary vector with components ξ^i and in the first case displace it parallel to itself to P_1, and from there to P_{12}, and in the second case first to P_2, and from there to P_{21}. Since P_{12} and P_{21} coincide, we can form the difference of these two vectors at this point and through this obviously obtain there a vector with the components

$$\Delta\xi^i = \delta d\xi^i - d\delta\xi^i.$$

From

$$d\xi^i = -d\gamma^i_k\,\xi^k = -\Gamma^i_{kl}dx_l\xi^k,$$

it follows that

$$\delta d\xi^i = -\frac{\partial \Gamma^i{}_{kl}}{\partial x_m} dx_l \delta x_m \xi^k - \Gamma^i{}_{kl} \delta dx_l \cdot \xi^k + d\gamma^i{}_r \delta \gamma^r{}_k \xi^k,$$

and because of the symmetry of δdx_l:

$$\Delta\xi^i = \left\{ \left(\frac{\partial \Gamma^i{}_{km}}{\partial x_l} - \frac{\partial \Gamma^i{}_{kl}}{\partial x_m} \right) dx_l \delta x_m + (d\gamma^i{}_r \delta \gamma^r{}_k - d\gamma^r{}_k \delta \gamma^i{}_r) \right\} \xi^k.$$

Therefore, we obtain

$$\Delta\xi^i = \Delta R^i{}_k \xi^k,$$

where the $\Delta R^i{}_k$ are linear forms of the two displacements d and δ, or rather of the area element spanned by them, independent of the vector ξ and with the components

$$\Delta x_{lm} = dx_l \delta x_m - dx_m \delta x_l,$$

$$\Delta R^i{}_k = R^i{}_{klm} dx_l \delta x_m = \frac{1}{2} R^i{}_{klm} \Delta x_{lm} \qquad (R^i{}_{kml} = -R^i{}_{klm}), \tag{4}$$

$$R^i{}_{klm} = \left(\frac{\partial \Gamma^i{}_{km}}{\partial x_l} - \frac{\partial \Gamma^i{}_{kl}}{\partial x_m} \right) + (\Gamma^i{}_{lr} \Gamma^r{}_{km} - \Gamma^i{}_{mr} \Gamma^r{}_{kl}). \tag{5}$$

If η_i are the components of an arbitrary force at P, then $\eta_i \Delta\xi^i$ is | an invariant; consequently, $R^i{}_{klm}$ are the components of a 4th order tensor at P, covariant in klm and contravariant in i, the *curvature*. That the curvature vanishes identically is the necessary and sufficient condition for the manifold to be Euclidean. In addition to the condition of "skew" symmetry given beside (4), the curvature components satisfy the condition of "cyclic" symmetry: [395]

$$R^i{}_{klm} + R^i{}_{lmk} + R^i{}_{mkl} = 0.$$

By its nature, the curvature at a point P is a linear map or transformation $\Delta \boldsymbol{P}$, which assigns to each vector ξ there another vector $\Delta\xi$; this transformation itself depends linearly on an element of area at P:

$$\Delta \boldsymbol{P} = \boldsymbol{P}_{ik} dx_i \delta x_k = \frac{1}{2} \boldsymbol{P}_{ik} \Delta x_{ik} \qquad (\boldsymbol{P}_{ki} = -\boldsymbol{P}_{ik}).$$

Accordingly, the curvature is best described as a "linear transformation-tensor of 2nd order."

In order to counter objections to the proof of the invariance of the curvature tensor, which could be raised against the above considerations involving infinitesimals, one uses a force field f_i, and forms the change $d(f_i \xi^i)$ of the invariant product $f_i \xi^i$ in such a way that under the infinitely small displacement d the vector ξ is displaced parallel to itself. Replacing in the expression obtained the infinitesimal displacement

dx with an arbitrary vector ρ at P, one obtains an invariant bilinear form of two arbitrary vectors ξ and ρ at P. From this one forms the change which corresponds to a second infinitely small displacement δ, by parallely taking along the vectors ξ and ρ, and replacing thereafter the second displacement by a vector σ at P. One obtains the form

$$\delta d(f_i\xi^i) = \delta df_i \cdot \xi^i + df_i\delta\xi^i + \delta f_i d\xi^i + f_i\delta d\xi^i.$$

Through the interchange of d and δ and subsequent subtraction, this yields, because of the symmetry of δdf_i, the invariant

$$\Delta(f_i\xi^i) = f_i\Delta\xi^i,$$

and thus the desired proof has been completed.

4. METRIC MANIFOLD (THE AETHER)

4.1 The Concept of The Metric Manifold

A manifold *carries at the point P a metric*, if the line elements at P can be compared with respect to their lengths. For this purpose, we assume the validity of the
[396] Pythagorean-Euclidean | laws in the infinitely small. Hence, to any two vectors ξ, η at P shall correspond a number $\xi \cdot \eta$, the *scalar product*, which is a symmetric bilinear form with respect to the two vectors. This bilinear form is certainly not absolute, but is only determined up to an arbitrary non-zero factor of proportionality. Hence, it is actually not the form $\xi \cdot \eta$, that is given but only the equation $\xi \cdot \eta = 0$; two vectors which satisfy this equation are called *perpendicular* to one another. We presuppose that this equation is non-degenerate, i.e. that the only vector at P, to which all vectors at P can be perpendicular is the 0 vector. We do not however presuppose that the associated quadratic form $\xi \cdot \xi$ is positive definite. If it has the index of inertia q, and if $n - q = p$, then we say in brief, the manifold at the point considered is $(p + q)$-dimensional. As a result of the arbitrary factor of proportionality, the two numbers p, q are only determined up to their order. We now assume that our manifold carries a metric [*Maßbestimmung*] at each point P. For the purpose of analytic representation, we consider (1) a definite coordinate system, and (2) the factor of proportionality appearing in the scalar product and which can be arbitrarily chosen at each point as fixed; with this, a "*frame of reference*"[9] for the analytic representation is obtained. If the vector ξ at the point P with the coordinates x_i has the components ξ^i, and η the components η^i, then one has

$$(\xi \cdot \eta) = \sum_{ik} g_{ik}\xi^i\eta^k \qquad (g_{ki} = g_{ik}),$$

9 I thus differentiate between "coordinate system" and "frame of reference."

where the coefficients g_{ik} are functions of the x_i. The g_{ik} should not only be continuous, but also be twice continuously differentiable. Since they are continuous and their determinant g by assumption does not vanish anywhere, the quadratic form $(\xi \cdot \xi)$ has the same index of inertia q at all points; therefore, we can describe the manifold in its entirety as $(p+q)$-dimensional. If we retain the coordinate system, but make a different choice for the undetermined factor of proportionality, then instead of the g_{ik} we obtain for the coefficients of the scalar product the quantities

$$g'_{ik} = \lambda \cdot g_{ik},$$

where λ is a nowhere vanishing continuous (and twice continuously differentiable) function of position.

According to the previous assumption, the manifold is only equipped with an *angle-measurement*; the geometry which is solely based on this, would be described as "*conformal geometry*"; it has, | as is well known, in the realm of two-dimensional [397]
manifolds ("Riemannian surfaces") experienced extensive development, because of its importance for complex function theory. If we make no further assumptions, then the individual points of the manifold remain completely isolated from one another with respect to metrical properties. The manifold becomes endowed with a metric connection from point to point, only when a *principle exists for the transfer of the unit of length from a point P to an infinitely close one*. Instead, Riemann made the much farther reaching assumption, that line elements can be compared not only at the same location, but that they can be compared as to their lengths at two finitely distant locations. *But the possibility of such a "non-local geometric" comparison definitely cannot be admitted in a purely infinitesimal geometry*. Riemann's assumption has also entered the Einsteinian world geometry of gravitation. Here, this inconsequence shall be removed.

Let P be a fixed point and P_* an infinitely close point obtained from P through the displacement with the components dx_i. We assume a definite frame of reference. In relation to the unit of length thus defined at P (as well as at all other points in the space), the square of the length of an arbitrary vector ξ at P is given by

$$\sum_{ik} g_{ik} \xi^i \xi^k.$$

Now, *if we transfer the unit of length chosen at P to P_**, which we presuppose as possible, the square of the length of an arbitrary vector ξ_* at P_* is given by

$$(1 + d\varphi) \sum_{ik} (g_{ik} + dg_{ik}) \xi^i_* \xi^k_*,$$

where $1 + d\varphi$ is a factor of proportionality deviating infinitesimally from 1; $d\varphi$ must be a homogeneous function of degree 1 of the differentials dx_i. Namely, if we transplant the unit of length chosen at P from point to point along a curve leading from P to a finitely distant point Q, then on the basis of the unit of length so

obtained at Q we obtain for the square of the length of an arbitrary vector at Q the expression $g_{ik}\xi^i\xi^k$, multiplied by the factor of proportionality which results from the product of the infinitely many individual factors of the form $1 + d\varphi$, which arise each time that we move from one point on the curve to the next.

$$\prod(1 + d\varphi) = \prod e^{d\varphi} = e^{\Sigma d\varphi} = e^{\int_P^Q d\varphi}.$$

[398] | In order that the integral appearing in the exponent makes sense, $d\varphi$ must be a function of the differentials of the kind asserted.

If one replaces g_{ik} by $g'_{ik} = \lambda g_{ik}$, then in place of $d\varphi$ a different quantity $d\varphi'$ will appear. If λ denotes the value of this factor at the point P, one must have

$$(1 + d\varphi')(g'_{ik} + dg_{ik}) = \lambda(1 + d\varphi)(g_{ik} + dg_{ik}),$$

and this yields

$$d\varphi' = d\varphi - \frac{d\lambda}{\lambda}. \tag{6}$$

Of the initially possible assumptions about $d\varphi$, that it is a linear differential form, or the root of a quadratic one, or the cubic root of a cubic one etc., only the first, as we can now see from (6), has an invariant meaning. We have thus arrived at the following result.

The metric of a manifold is based on a quadratic and on a linear differential form

$$ds^2 = g_{ik}dx_i dx_k \qquad \text{and} \qquad d\varphi = \varphi_i dx_i. \tag{7}$$

However, conversely these forms are not absolutely determined by the metric, but each pair of forms ds'^2 and $d\varphi'$, which arise from (7) according to the equations

$$ds'^2 = \lambda \cdot ds^2, \qquad d\varphi' = d\varphi - \frac{d\lambda}{\lambda} \tag{8}$$

is equivalent to the first pair in the sense that both express the same metric. In this λ is an arbitrary, nowhere vanishing continuous (more precisely: twice continuously differentiable) *function of position*. Into all quantities or relations which represent metric relations analytically, the functions g_{ik}, φ_i must thus enter in such a way that invariance holds (1) with respect to an arbitrary coordinate transformation ("coordinate-invariant"), and (2) with respect to the replacement of (7) by (8) ("measure-invariance").

$$\frac{d\lambda}{\lambda} = d \lg \lambda$$

is a total differential. Hence, whereas in the quadratic form ds^2, a factor of proportionality remains arbitrary at each location, the indeterminacy of $d\varphi$ consists of an additive total differential.

A metric manifold we describe physically as a world filled with *aether*. The particular metric existing in the manifold represents a particular state of the world filling aether. This state is thus to be described relative to a frame of reference through the specification (arithmetic construction) of the functions g_{ik}, φ_i. |

From (6) it follows that the linear tensor of 2nd order with the components [399]

$$F_{ik} = \frac{\partial \varphi_i}{\partial x_k} - \frac{\partial \varphi_k}{\partial x_i}$$

is uniquely determined by the metric of the manifold; I call it the *metric vortex*. It is the same, I believe, as what in physics one calls the *electromagnetic field*. It satisfies the "first system of Maxwell's equation"

$$\frac{\partial F_{kl}}{\partial x_i} + \frac{\partial F_{li}}{\partial x_k} + \frac{\partial F_{ik}}{\partial x_l} = 0.$$

Its vanishing is the necessary and sufficient condition for the transfer of length to be integrable, i.e., for those conditions which Riemann placed at the foundations of metric geometry to prevail. We understand from this how Einstein through his world geometry, which mathematically follows Riemann, could only account for gravitation but not for the electromagnetic phenomena.

4.2 Affine Connection of a Metric Manifold

In a metric space, in place of the requirement A imposed on the concept of parallel displacement in §3, I., we have the more specific one

A: that the parallel displacement of all vectors at a point P to an infinitely close point P', must not only be an affine but also a congruent transfer of the totality of these vectors*.

Using the previous notation, this requirement yields the equation

$$(1 + d\varphi)(g_{ik} + dg_{ik})(\xi^i + d\xi^i)(\xi^k + d\xi^k) = g_{ik}\xi^i\xi^k. \tag{9}$$

For all quantities a^i, which carry an upper index (i), we define the "lowering" of the index through the equations

$$a_i = \sum_k g_{ik} a^k.$$

(and the reverse process of raising an index through the inverse equations). Using this symbolism, for (9) we can write

$$(g_{ik}\xi^i\xi^k)d\varphi + \xi^i\xi^k dg_{ik} + 2\xi_i d\xi^i = 0.$$

The last term is

$$= -2\xi_i\xi^k d\gamma^i{}_k = -2\xi^i\xi^k d\gamma_{ik} = -\xi^i\xi^k(d\gamma_{ik} + d\gamma_{ki});$$

[400] | and therefore

$$d\gamma_{ik} + d\gamma_{ki} = dg_{ik} + g_{ik}d\varphi. \tag{10}$$

This equation can certainly be satisfied only if $d\varphi$ is a linear differential form; an assumption to which we were already driven above as the only reasonable one. From (10) or

$$\Gamma_{i,kr} + \Gamma_{k,ir} = \frac{\partial g_{ik}}{\partial x_r} + g_{ik}\varphi_r \tag{10*}$$

follows, as a consequence of the symmetry property $\Gamma_{r,ik} = \Gamma_{r,ki}$:

$$\Gamma_{r,ik} = \frac{1}{2}\left(\frac{\partial g_{ir}}{\partial x_k} + \frac{\partial g_{kr}}{\partial x_i} - \frac{\partial g_{ik}}{\partial x_r}\right) + \frac{1}{2}(g_{ir}\varphi_k + g_{kr}\varphi_i - g_{ik}\varphi_r); \quad (\Gamma_{r,ik} = g_{rs}\Gamma^s{}_{ik}). \tag{11}$$

It turns out that on a metric manifold the concept of the infinitesimal parallel displacement of a vector is uniquely determined through the requirements put forward.[10] I consider this as the *fundamental fact of infinitesimal geometry*, that with the metric also the affine connection of a manifold is given, that *the principle of transfer of length inherently carries with it that of transfer of direction*, or expressed physically, *that the state of the aether determines the gravitational field*.

If the quadratic form $g_{ik}dx_i dx_k$ is indefinite, then among the geodesic lines, the null lines are distinguished as those along which the form vanishes. They depend only on the ratios of the g_{ik}, but not at all on the φ_i, they are thus structures of conformal geometry.[11]

We had imposed certain axiomatic requirements on the concept of parallel transport and shown that they can be satisfied on a metric manifold in one and only one way. However, it is also possible to define that concept explicitly in a simple manner. If P is a point in our metric manifold, then we call a frame of reference *geodesic* in P, if upon its use the φ_i vanish at P and the g_{ik} assume stationary values:

$$\varphi_i = 0, \qquad \frac{\partial g_{ik}}{\partial x_r} = 0.$$

[401] | D. *For each point P there exist geodesic frames of reference. If ξ is a given vector at P, and P' is an infinitely close point to P, then we understand by the vector which arises from x through parallel transport to P' that vector at P', which has the same components as ξ in the geodesic coordinate system belonging to P. This definition is independent of the choice of the geodesic frame of reference.*

---

10 See also Hessenberg, "Vektorielle Begründung der Differentialgeometrie," *Math. Ann.* vol. 78 (1917), p. 187–217, especially p. 208.

11 With this comment, I would like to correct a mistake on page 183 of my book *Raum, Zeit, Materie*.

It is not difficult to demonstrate the assertion contained in this explanation independently of the train of thought followed here through direct calculation, and to show by the same means that the process of parallel transport so defined is, in an arbitrary coordinate system, described by the equation

$$d\xi^r = -\Gamma^r_{ik}\xi^i dx_k \tag{12}$$

with the coefficients Γ to be taken from (11).[12] But here, where the invariant meaning of equation (12) is already established, we conclude more simply as follows. According to (11), the Γ^r_{ik} vanish in a geodesic frame of reference and the equations (12) reduce to $d\xi^r = 0$. Hence, the concept of parallel transfer that we derived from the axiomatic requirements agrees with the one defined in D. Only the existence of a geodesic frame of reference is left to be shown. For this purpose, we choose a coordinate system x_i, geodesic at P, having the point P as its origin $(x_i = 0)$. If the unit of length at P and in its vicinity is for the time being chosen arbitrarily, and if furthermore the φ_i denote the value of these quantities at P, then one only needs to complete the transition from (7) to (8) with

$$\lambda = e^{\Sigma_i \varphi_i x_i},$$

in order to obtain that, besides the Γ^i_{rs}, the φ_i also vanish at P. From this then follows—see (10*)—the geodesic nature of the frame of reference so obtained. The coordinates of a frame of reference geodesic at P are in the immediate vicinity of P determined up to terms of 3rd order, leaving aside linear transformation, and the unit of length up to terms of 2nd order, leaving aside the addition of a constant factor.

12 In this one could follow the approach I have taken in RZM, §14.

ELIE CARTAN

THE DYNAMICS OF CONTINUOUS MEDIA AND THE NOTION OF AN AFFINE CONNECTION ON SPACE-TIME

Originally published as chapter 1 of "Sur les variétés à connexion affine et la théorie de la relativité généralisée" in Annales Scientifiques de L'Ecole Normale Supérieure (1923): 325–412. Translation, by Anne Magnon and Abhay Ashtekar, taken from "On Manifolds With An Affine Connection And The Theory Of General Relativity," (Napoli: Bibliopolis, 1986), p. 31–55.

PRINCIPLE OF INERTIA AND NEWTONIAN GRAVITY

1. Newton's foundation of classical mechanics rests on the concepts of absolute time and absolute space. Thus, analytically, any event can be labelled in time and space provided a choice is made of an origin, a unit of time, and a frame of spatial coordinates. For example, the frame might originate at the center of mass of the solar system and its axes might point towards fixed stars. Of course, any other frame which is invariantly related to this one would be also admissible. As is well known, the laws of mechanics remain unaffected if the frame of spatial coordinates is made to undergo a rectilinear, uniform translation with respect to Newton's absolute space, keeping the absolute time undisturbed. One is thus led to the notion of *Galilean frames of reference*.

The principle of inertia may be stated as follows: *in absence of interactions with other bodies, the velocity of a point mass remains constant in direction and magnitude in any Galilean frame*. The fact that the validity of this principle in one Galilean frame implies its validity in any other Galilean frame follows immediately from the transformation laws governing these frames. Let us label a point in space by arbitrary Cartesian coordinates,[1] not necessarily orthogonal. Then the transformation laws are as follows:

1 In the passage from orthogonal to general coordinates, the laws of classical mechanics retain their form and the formulas of *theoretical mechanics* remain unchanged.

Jürgen Renn (ed.). *The Genesis of General Relativity,* Vol. 4
Gravitation in the Twilight of Classical Physics: The Promise of Mathematics.
© 2007 Springer.

$$\left.\begin{aligned} x' &= a_1x + b_1y + c_1z + g_1t + h_1 \\ y' &= a_2x + b_2y + c_2z + g_2t + h_2 \\ z' &= a_3x + b_3y + c_3z + g_3t + h_3 \\ t' &= t + h, \end{aligned}\right\}$$

where the coefficients are constants. One can now give an alternate formulation of the
[32] principle of inertia. Two reference frames will be | said to be *equivalent* provided they are equivalent in the usual geometric sense and motionless relative to each other. Thus, the coordinate transformation between equivalent frames is given by:

$$\left.\begin{aligned} x' &= x + h_1, \\ y' &= y + h_2, \\ z' &= z + h_3, \\ t' &= t + h. \end{aligned}\right\}$$

Now, consider a moving point mass and attach to it, at each instant of time, a Galilean frame which has that point as its origin.[2] Then the principle of inertia can be stated as follows: *if a system of equivalent Galilean frames is attached to a moving point mass as above, then, at any instant of time, the velocity of the point mass in the Galilean frame corresponding to that instant is constant if the point mass is not subject to interaction with other bodies*.

2. Clearly, the structure of mechanics is based on two notions:

(i)The notion of a Galilean frame (which enables one to define the velocity of a moving point mass);

(ii)The notion of equivalent Galilean frames (which enables one to state the principle of inertia).

It is important to note the advantage of the second formulation of the principle of inertia: in essence, *it uses the notion of equivalent Galilean frames only for those frames whose origins are infinitesimally close*. All generalizations of classical or relativistic mechanics retain the notion of Galilean frames; it is the notion of equivalent frames that has undergone modifications.

Let us continue to use the framework of classical mechanics with the notion of absolute time (as measured by a unit which is fixed once and for all). We shall see that a modification of the notion of equivalent frames will enable us to extend the principle of inertia so that it incorporates not only isolated point masses but also *point*

2 That is, the point is the origin of the coordinate axes, and the instant, at which one examines it, is taken to be the origin of time.

masses placed in a gravitational field. Let us fix a Galilean frame and denote by T_0 the corresponding spatial triad of coordinates. Next, let us introduce a field of forces analogous to a gravitational field, i.e., an acceleration field (X, Y, Z). Then, if the velocity of a point mass w.r.t. the fixed Galilean frame is given by

$$u, \quad v, \quad w$$

at time t, at time $t + dt$ the velocity will be

$$u + Xdt, \quad v + Ydt, \quad w + Zdt.$$

| At time t, let us attach to the point mass a triad T which is equivalent to T_0 in the usual geometrical sense. Similarly, at time $t + dt$, let us attach a triad T'. These triads will define Galilean frames only if one specifies that they are in a rectilinear, uniform motion w.r.t. T_0. When this is done, the resulting Galilean frames will have origins [33]

$$x, \quad y, \quad z, \quad t \text{ and } x + dx, \quad y + dy, \quad z + dz, \quad t + dt$$

respectively. Denote by

$$a, \quad b, \quad c, \text{ and } a', \quad b', \quad c'$$

the translation velocities of T and T' w.r.t. T_0. Then the velocity of the point mass in the Galilean frame attached to it at time t has components

$$u - a, \quad v - b, \quad w - c,$$

and in the Galilean frame attached at time $t + dt$,

$$u + Xdt - a', \quad v + Ydt - b', \quad w + Zdt - c'.$$

Thus the components will not have changed if

$$a' - a = Xdt, \quad b' - b = Ydt, \text{ and } \quad c' - c = Zdt.$$

Consequently, *the motion of an arbitrary point mass which is placed in the field of forces described above will satisfy the principle of inertia provided two Galilean frames with infinitesimally close origins,*

$$x, \quad y, \quad z, \quad t;$$

$$x + dx, \quad y + dy, \quad z + dz, \quad t + dt,$$

are considered as equivalent if their triads T and T' are equivalent in the usual geometrical sense, and if T is in a rectilinear, uniform translational motion w.r.t. T' with velocity (Xdt, Ydt, Zdt). Note that, again, we have used mutual relations only between the infinitesimally close frames of reference.

3. One can express the same ideas in a way which is perhaps more intuitive, and which has the advantage of being closer to the viewpoint adopted by Einstein than the
[34] point of departure for his theory of | gravitation. Consider a point particle moving in the field of forces discussed above and attach to it a spatial triad T originating at the point and carried by it in a *translational* motion. At each instant of time, introduce a Galilean reference frame consisting of a triad $\overline{T}$ which coincides with T at the instant considered and which is in a *rectilinear*, *uniform* translational motion, with the velocity which the particle has at that precise instant.[3] Obviously, the velocity of the particle w.r.t. these frames is zero. Thus, the motion of the point mass agrees with the principle of inertia (constancy of velocity) if the successive Galilean reference frames defined by the triads $\overline{T}$ are considered as equivalent, *step by step*. Clearly, the constant velocity of the triad $\overline{T'}$, corresponding to time $t + dt$, w.r.t. the triad $\overline{T}$, corresponding to time t, is given by

$$X dt, \quad Y dt, \quad Z dt.$$

Consider for example the uniform field due to earth's gravity and assume for a moment that one can neglect the motion of the earth w.r.t. the absolute space. Choose the z-axis along the vertical upwards direction. Two triads, T and T', associated with instants of time t and $t + dt$, will be said to define two equivalent Galilean frames of reference if T' has a constant vertical velocity $g dt$ w.r.t. T. Thus, one can attach a Galilean frame to each event (x, y, z, t) of space-time in such a way that *all the resulting frames are equivalent*; given the frame associated with a particular event (x_0, y_0, z_0, t_0), all others will be completely determined. In a general case, however, such a situation does not occur: Whether two frames are equivalent or not will depend on the space-time paths connecting their origins, since the notion of equivalence itself has been introduced via a step by step procedure. We shall return to this fundamental issue later on.

4. Let us say that the conditions determining the equivalence of two Galilean frames with infinitesimal close origins define the *geometrical*[4] properties of the space-time. Thus, the gravitational phenomena are shifted from the domain of physics to that of geometry[5] and the components X, Y and Z of the gravitational field capture the basic geometrical structure of space-time.[6] The relations

$$\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z} = 0, \qquad \frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x} = 0, \qquad \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y} = 0 \tag{1}$$

---

3 Actually, for the purposes for which the triad $\overline{T}$ has been used, one could have replaced it by T itself; thus, as far as the velocity of a point at instant t is concerned, T plays the role of a Galilean triad.

4 Actually, these properties are geometrical as well as kinematical.

5 This is essentially another way of stating the equality of inertial and gravitational masses, or, the fact that the gravitational field is kinematical (a field of accelerations) rather than dynamical (a field of forces).

which hold in *orthogonal* coordinates express the properties of this | structure. [35]
Finally, the fundamental Poisson equation,

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = -4\pi\rho, \tag{2}$$

which, together with the above relations, yields a complete formulation of the laws of Newtonian gravitation,[7] shows that *the matter density of a continuous medium is the physical manifestation of a local geometrical property of space-time*. Thus, we recover some features of Einstein's theory of gravity within the framework of classical mechanics itself. The only essential difference is the lack of relation between gravitational and electromagnetic phenomena. But we have already recovered the structure which intertwines space-time, geometry and matter.

5. All these considerations call for further remarks. Does the reduction of gravitation to geometry occur only for a specific definition of the equivalence of two infinitesimally close Galilean frames? We shall examine this question in detail later. For the time being, let me just say that the answer is in the negative. Let us consider two Galilean frames with infinitesimally close origins which are equivalent in the sense of section 3 above. Thus, the corresponding triads T and T', with origins M and M' are parallel, T' undergoing a uniform rectilinear translation w.r.t. (T). Let us now replace T' by T'', a triad which is fixed w.r.t. T', has M' as its origin, and is obtained from (T) by a helicoidal displacement along the axis MM', the sense and the magnitude of the displacement being fixed once and for all. Consider a point mass which if freely falling in a gravitational field such that it finds itself at M at time t and at M' at time $t + dt$. Since the velocity of this point mass is almost colinear with MM' at time $t + dt$, it will have the same components w.r.t. the frame S' defined by T' as those w.r.t. the frame S'' defined by T''. Thus, if the motion of the point mass obeys the principle of inertia when S and S' are equivalent, *it will continue to obey this principle with the modified definition of equivalence*. This example leads us to the following conclusion: *As far as the dynamics of a point mass is concerned, there exists an infinite number of definitions of equivalence of Galilean frames whose origins are infinitesimally close*.

6. One might expect that these conclusions would have to be modified for the dynamics of material *systems* since the dynamics of a point mass neglects the important issue of rotation. Consider a small spherical ball undergoing an absolute, uniform

---

6 In fact, as we shall see later, this structure requires the functions X, Y, Z only to be defined up to arbitrary additive constants. This is because mechanical experiments performed inside a system which is embedded in a *uniform* gravitational field cannot detect this field. In particular, if one assumes that the gravitational field due to distant stars is *uniform over the solar system,* the laws of celestial Mechanics governing the motion of sun and its planets remain unchanged.

7 In addition, we must assume that the functions X, Y, Z vanish at infinity.

rotation. The axis of rotation, along which its angular momentum points, should be |
[36] considered as remaining equivalent to itself. Thus, our first convention by which spatial directions remain parallel to themselves in the usual sense, appears to be the only one that is permissible. However, it is simply too early to draw such a conclusion. In fact, we shall see later on that this conclusion is premature and *the high degree of indeterminacy in the notion of equivalence persists in its entirety when one deals with the laws of dynamics of material systems*.[8] However, to investigate this issue in a fruitful way, it is important to note that the new viewpoint which we have now adopted requires that the laws of mechanics should be formulated only *locally*. In other words, we must go back to mechanics of continuous media. Indeed, we do not have the notion of equivalence of two frames unless their origins are infinitesimally close.

In order to facilitate the transition from Newtonian to relativistic mechanics, we shall now formulate the equations of mechanics of continuous media using a 4-dimensional manifold as the model for space-time.

FOUR DIMENSIONAL SPACE-TIME AND CLASSICAL DYNAMICS OF CONTINUOUS MEDIA

7. Let us adopt the viewpoint of classical mechanics. Space-time or the universe will be represented by an *affine manifold*. By this, we mean the following. Let us call a *space-time vector* a set consisting of two events (each located in time and space) one of which is the origin of the vector and, the other, the extremity. In a Galilean frame the components of a space-time vector are the four numbers

$$t' - t, \quad x' - x, \quad y' - y, \quad z' - z,$$

obtained by subtracting the coordinates of the origin from those of the extremity. If the components of two space-time vectors are identical in one Galilean frame, they are identical in all Galilean frames. Thus, here we have a property of space-time vectors which is independent of the Galilean frame used to represent these vectors mathematically. Vectors which have this property will be said to be *equivalent*. It is clear that if two space-time vectors are equivalent to a third, they are equivalent to each other. It is the existence of this notion of equivalence among space-time vectors that we express when we say that space-time has an affine structure.

Of the four numbers,

$$t' - t, \quad x' - x, \quad y' - y, \quad z' - z$$

[37] | which mathematically represent a space-time vector, the first will be referred to as the *time component* and the remaining three will be called *space components*. Note that *the time component is independent of the choice of reference frame*. The situation is different for the spatial vector whose components in the coordinate triad T defin-

8 Except for one possible restriction; see section no. 16.

ing the given Galilean frame are $x' - x$, $y' - y$ and $z' - z$; the spatial vector depends not only on the given space-time vector and the triad T but also on the velocity of this triad w.r.t. the absolute space.

Let us now consider a point mass moving w.r.t. a Galilean frame. In this frame, the components of the space-time vector joining the position of this point mass at time t to that at time $t + dt$ are:

$$dt, \quad dx, \quad dy, \quad dz.$$

The vector itself does not depend on the Galilean frame. The same remarks hold for the space-time vector

$$1, \quad \frac{dx}{dt}, \quad \frac{dy}{dt}, \quad \frac{dz}{dt}$$

which is obtained by dividing the first space-time vector by dt. Finally, if we denote the mass of the particle by m, the space-time vector

$$m, \quad m\frac{dx}{dt}, \quad m\frac{dy}{dt}, \quad m\frac{dz}{dt}$$

is itself again independent of one's choice of the frame of reference. This is the *energy-momentum vector*. While its time component, the mass, is independent of the choice of the frame of reference, its space-component, the momentum, does depend on this choice.

We can now state the fundamental principle of particle dynamics: *The time-derivative of the energy-momentum space-time vector is equal to the spatial force vector.* This statement contains both the principle of conservation of mass and the law relating force and acceleration.

8. Let us now consider a continuous medium equipped with a given Galilean frame. Fix a 3-dimensional volume of space-time. As I have shown elsewhere,[9] the total mass contained in this volume is given by the integral

$$\iiint \rho dxdydz - \rho u dydzdt - \rho v dzdxdt - \rho w dxdydt$$

where ρ denotes the density and u, v, w denote the components of | the velocity of [38]
each element of matter. Let us first assume that the matter is *free of pressure as well as stress*. Then the x component of momentum of the same volume is given by the integral

$$\iiint \rho u dxdydz - \rho u^2 dydzdt - \rho uv dzdxdt - \rho uw dxdydt.$$

The y and the z components can be expressed similarly. Let us denote by

---

9 E. Cartan, *Leçons sur les Invariants intégraux*, Paris, Hermann, 1922, p. 35-37. [Translator's note: The integral is just $\iiint p^a_{\varepsilon abcd} dx^b \wedge dx^c \wedge dx^d$, where $p^a \equiv (\rho, \rho u, \rho v, \rho w)$ is the 4-momentum density.]

$$\Pi,\quad \Pi_x,\quad \Pi_y,\quad \Pi_z$$

the integrands of these integrals. These are the four components of the energy-momentum vector of an element of matter in the medium. Finally, let us denote by $X, Y, Z,$ the components of the force per unit volume.

To obtain the equations of mechanics of continuous media, we proceed as follows. Consider a 4-dimensional domain of space-time and decompose it into *world tubes* formed by elements of the matter under consideration, taken between time t_1 and t_2. Thus the boundary of this domain consists of elements of matter at the extremities t_1 and t_2 of the time interval. Now, the geometrical difference between the energy-momentum vectors of a matter element evaluated at time $t_1,$ when the element enters the domain, and at time $t_2,$ when it leaves, is given by a spatial vector with components

$$\int_{t_1}^{t_2}(X dx dy dz)dt, \qquad \int_{t_1}^{t_2}(Y dx dy dz)dt, \qquad \int_{t_1}^{t_2}(Z dx dy dz)dt.$$

In other words, *the integral of the "energy-momentum 4-vector" over the boundary of the domain is equal to the four dimensional integral of the "force 3-vector" over the domain itself*. This statement finds its expression in the following formulas:

$$\left.\begin{aligned} \Pi' &= 0 \\ \Pi'_x &= X dt dx dy dz \\ \Pi'_y &= Y dt dx dy dz \\ \Pi'_z &= Z dt dx dy dz, \end{aligned}\right\} \tag{5}$$

where

$$\Pi' = \left[\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z}\right] dt dx dy dz$$

$$\Pi'_x = \left[\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z}\right] dt dx dy dz$$

$$\Pi'_y = \left[\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho uv)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho vw)}{\partial z}\right] dt dx dy dz$$

$$\Pi'_z = \left[\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho wu)}{\partial x} + \frac{\partial(\rho wv)}{\partial y} + \frac{\partial(\rho w^2)}{\partial z}\right] dt dx dy dz.$$

[39] | After some calculations and simplifications these equations yield the familiar ones:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0$$

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) = X$$

$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) = Y$$

$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) = Z.$$

Let us call *exterior derivative*[10] the operation which enables one to convert an integral over a closed p-dimensional manifold to the integral over the $(p+1)$-dimensional manifold enclosed by the p-manifold. Then the fundamental principle of mechanics of continuous media can be stated as follows: *The exterior derivative of the energy-momentum field is equal to the product of dt with the force field*.

9. In the above discussion, we had assumed the absence of pressure as well as stress. However, the general case can be reduced to the one discussed above by defining the (generalized) momentum of an element of matter to be the vector whose components are obtained by adding the following quantities to the components introduced previously:

$$-p_{xx}dydzdt - p_{xy}dzdxdt - p_{xz}dxdydt$$
$$-p_{yx}dydzdt - p_{yy}dzdxdt - p_{yz}dxdydt$$
$$-p_{zx}dydzdt - p_{zy}dzdxdt - p_{zz}dxdydt.$$

In the kinetic theory of gases, one can in effect consider pressure to be the flux of momentum resulting from irregularities in the molecular velocities. On the other hand, the quantities u, v, w introduced previously represent only an *average* velocity.

The usual equations of mechanics of continuous media can be | now recovered by [40]
expanding equations (5):

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0,$$

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) + \frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{xy}}{\partial y} + \frac{\partial p_{xz}}{\partial z} = X$$

$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) + \frac{\partial p_{yx}}{\partial x} + \frac{\partial p_{yy}}{\partial y} + \frac{\partial p_{yz}}{\partial z} = Y$$

$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) + \frac{\partial p_{zx}}{\partial x} + \frac{\partial p_{zy}}{\partial y} + \frac{\partial p_{zz}}{\partial z} = Z.$$

10 See E. Cartan, *Leçons sur les Invariants intégraux*, Chapter VII, p. 65.

10. However, these equations are not complete. In effect, this is because we have not taken into account the theorem[11] of angular momentum which may be expressed in the present framework as follows:

$$\iiint y\Pi_z - z\Pi_y = \iiiint (yZ - zY)dtdxdydz$$

$$\iiint z\Pi_x - x\Pi_z = \iiiint (zX - xZ)dtdxdydz$$

$$\iiint x\Pi_y - y\Pi_x = \iiiint (xY - yX)dtdxdydz,$$

where the integrals on the right-hand side are taken over an arbitrary volume element of space-time and those on the left hand side, on the 3-dimensional boundary of this volume. These equations yield:

$$[dy\Pi_z] - [dz\Pi_y] = 0$$
$$[dz\Pi_x] - [dx\Pi_z] = 0$$
$$[dx\Pi_y] - [dy\Pi_x] = 0,$$

and they are satisfied trivially in absence of pressure. In the general case, they imply:

$$p_{zy} - p_{yz} = 0$$
$$p_{xz} - p_{zx} = 0$$
$$p_{yx} - p_{xy} = 0.$$

11. One can express the previous results using a simple vectorial notation. Let e_0, e_1, e_2, e_3 denote the 4-vectors whose components are, respectively, $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$. The last three are spatial
[41] vectors. In this notation, the energy-momentum | of a particle of mass m is given by

$$m\left(e_0 + \frac{dx}{dt}e_1 + \frac{dy}{dt}e_2 + \frac{dz}{dt}e_3\right).$$

Let us now denote by m the space-time point (t, x, y, z). Then the derivative dm/dt of this point w.r.t. time is a space-time vector with components

$$\left(1, \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right).$$

Thus, the energy-momentum of the particle is given by

11 Note that the analytical formulation of this theorem does not require the restriction to rectangular axes.

$$m\frac{d\mathrm{m}}{dt}.$$

The points and the (free) vectors are "geometric forms" of order one. One can also consider geometric forms of second order which represent systems of *sliding vectors*. We shall denote by $[\mathrm{mm}']$ the sliding vector whose origin lies at the space-time point m and whose extremity is at the space-time point m'. This sliding vector has ten plückerian coordinates which are the 2×2 determinants constructed from the tableau

$$\begin{matrix} 1, & t, & x, & y, & z \\ 1, & t', & x', & y', & z'; \end{matrix}$$

clearly,

$$[\mathrm{mm}'] = -[\mathrm{m}'\mathrm{m}].$$

Similarly, we shall denote by $[\mathrm{me}]$ the sliding vector obtained from the vector which originates at the space-time point m and which is parallel to a given vector e. The plückerian coordinates of this vector are obtained from the tableau

$$\begin{matrix} 1, & t, & x, & y, & z, \\ 0, & \theta, & \xi, & \eta, & \zeta, \end{matrix}$$

where the second line contains the components of the vector e. Finally, | let us denote [42]
by $[\mathrm{ee}']$ the bivector whose ten coordinates are obtained from the tableau

$$\begin{matrix} 0, & \theta, & \xi, & \eta, & \zeta, \\ 0, & \theta', & \xi', & \eta', & \zeta', \end{matrix}$$

of the components of the two free vectors e and e'.

In each of these cases, the sliding vector or the bivector under consideration may be viewed as the (exterior) product of the two factors each of which is a first order form (a point or a free vector). The product is distributive and antisymmetric.

12. The sliding vector whose origin lies at the space-time point m representing the position of a point particle at a given instant of time and which carries the energy-momentum of the particle can be expressed as

$$m\left[\mathrm{m}\frac{d\mathrm{m}}{dt}\right].$$

Therefore, the equation

$$\frac{d}{dt}\left\{m\left[\mathrm{m}\frac{d\mathrm{m}}{dt}\right]\right\} = [\mathrm{F}],$$

where [F] is the "force" sliding vector, contains, at once, the fundamental principle of dynamics and the theorem of angular momentum. Indeed, it contains the ten equations

$$\frac{dm}{dt} = 0, \qquad \frac{d}{dt}\left(m\frac{dx}{dt}\right) = X,$$

$$\frac{d}{dt}\left(m\frac{dy}{dt}\right) = Y, \qquad \frac{d}{dt}\left(m\frac{dz}{dt}\right) = Z,$$

$$\frac{d}{dt}\left(mt\frac{dx}{dt} - mx\right) = tX, \qquad \frac{d}{dt}\left(mt\frac{dy}{dt} - my\right) = tY,$$

$$\frac{d}{dt}\left(mt\frac{dz}{dt} - mz\right) = tZ, \qquad \frac{d}{dt}\left(my\frac{dz}{dt} - mz\frac{dz}{dt}\right) = yZ - zY,$$

$$\frac{d}{dt}\left(mz\frac{dx}{dt} - mx\frac{dz}{dt}\right) = zX - xZ, \frac{d}{dt}\left(mx\frac{dy}{dt} - my\frac{dx}{dt}\right) = xY - yX.$$

[43] | 13. Let us now return to the mechanics of continuous media. Denote the energy-momentum by a sliding vector G and the force per unit element of a 3-dimensional volume by a sliding vector F. *Then the equations of mechanics are succintly captured in the single formula*

$$\mathrm{G}' = [dt\mathrm{F}]. \tag{6}$$

Note that

$$\mathrm{G} = [m\mathrm{e}_0]\Pi + [m\mathrm{e}_1]\Pi_x + [m\mathrm{e}_2]\Pi_y + [m\mathrm{e}_3]\Pi_z,$$

and

$$\mathrm{F} = [m\mathrm{e}_1]X dxdydz + [m\mathrm{e}_2]Y dxdydz + [m\mathrm{e}_3]Z dxdydz.$$

The equation

$$d\mathrm{m} = \mathrm{e}_0 dt + \mathrm{e}_1 dx + \mathrm{e}_2 dy + \mathrm{e}_3 dz$$

yields

$$\begin{aligned}\mathrm{G}' = {} & [m\mathrm{e}_0]\Pi' + [m\mathrm{e}_1]\Pi'_x + [m\mathrm{e}_2]\Pi'_y + [m\mathrm{e}_3]\Pi'_z \\ & + [\mathrm{e}_0\mathrm{e}_1][dt\Pi_x - dx\Pi] + [\mathrm{e}_0\mathrm{e}_2][dt\Pi_y - dy\Pi] + [\mathrm{e}_0\mathrm{e}_3][dt\Pi_z - dz\Pi] \\ & + [\mathrm{e}_2\mathrm{e}_3][dy\Pi_z - dz\Pi_y] + [\mathrm{e}_3\mathrm{e}_1][dz\Pi_x - dx\Pi_z] + [\mathrm{e}_1\mathrm{e}_2][dx\Pi_y - dy\Pi_x].\end{aligned}$$

It is easy to verify that the coefficients of $[\mathrm{e}_0\mathrm{e}_1]$, $[\mathrm{e}_0\mathrm{e}_2]$ and $[\mathrm{e}_0\mathrm{e}_3]$ vanish identically. If the elements of matter are subject *to a torque*, *in addition to the force*, one has simply to add terms of the following form to the expression of the force:

$$[\mathrm{e}_2\mathrm{e}_3]\mathrm{L}dxdydz + [\mathrm{e}_3\mathrm{e}_1]\mathrm{M}dxdydz + [\mathrm{e}_1\mathrm{e}_2]\mathrm{N}dxdydz.$$

This implies

$$p_{zy} - p_{yz} = \mathrm{L}, \qquad p_{xz} - p_{zx} = \mathrm{M}, \qquad p_{yx} - p_{xy} = \mathrm{N}.$$

Then the fundamental equation (6) continues to hold. Clearly, the basic equation of dynamics can be recovered from equation (6) under the assumption that the matter is contained in a very small spatial volume: in this approximation, one obtains

$$d\mathrm{G} = dt\mathrm{F}.$$

| 14. Equation (6) will enable us to obtain easily the equations of mechanics of [44] continuous media. Let us attach a variable Galilean frame to each point of space-time. Denote by $\mathrm{e}_0, \mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ the free space-time vectors which define the Galilean frame attached to the point m. As we move from a point m to an infinitesimally nearby point m', these vectors will change. However, the time component of e_0 will be always equal to 1, and those of $\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ will always vanish. One will therefore have the following formulae:[12]

$$\begin{aligned} d\mathrm{e}_0 &= \omega_0^1 \mathrm{e}_1 + \omega_0^2 \mathrm{e}_2 + \omega_0^3 \mathrm{e}_3 \\ d\mathrm{e}_1 &= \omega_1^1 \mathrm{e}_1 + \omega_1^2 \mathrm{e}_2 + \omega_1^3 \mathrm{e}_3 \\ d\mathrm{e}_2 &= \omega_2^1 \mathrm{e}_1 + \omega_2^2 \mathrm{e}_2 + \omega_2^3 \mathrm{e}_3 \\ d\mathrm{e}_3 &= \omega_3^1 \mathrm{e}_1 + \omega_3^2 \mathrm{e}_2 + \omega_3^3 \mathrm{e}_3 \end{aligned} \tag{7}$$

where ω_i^j are linear combinations of the differentials of the four functions which label space-time points. Let us denote by

$$d\mathrm{m} = \omega^0 \mathrm{e}_0 + \omega^1 \mathrm{e}_1 + \omega^2 \mathrm{e}_2 + \omega^3 \mathrm{e}_3 \tag{8}$$

the free vector joining m and m'. Thus ω^0 is simply the infinitesimal time interval between m and m'. Now, one can again obtain the following expressions:

$$\mathrm{G} = [\mathrm{me}_0]\Pi + [\mathrm{me}_1]\Pi_x + [\mathrm{me}_2]\Pi_y + [\mathrm{me}_3]\Pi_z,$$

and

$$\mathrm{F} = [\mathrm{me}_1]X\omega^1\omega^2\omega^3 + [\mathrm{me}_2]Y\omega^1\omega^2\omega^3 + [\mathrm{me}_3]Z\omega^1\omega^2\omega^3.$$

It is only the expression of G' that becomes more complicated because the free vectors $\mathrm{e}_0, \mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ are no longer fixed. One has:

12 As in section no. 1, the axes of coordinate frames are not necessarily orthogonal here.

$$
\begin{aligned}
G' = {} & [me_0]\Pi' + [me_1][\Pi'_x + \omega_0^1\Pi + \omega_1^1\Pi_x + \omega_2^1\Pi_y + \omega_3^1\Pi_z] \\
& + [me_2][\Pi'_y + \omega_0^2\Pi + \omega_1^2\Pi_x + \omega_2^2\Pi_y + \omega_3^2\Pi_z] \\
& + [me_3][\Pi'_z + \omega_0^3\Pi + \omega_1^3\Pi_x + \omega_2^3\Pi_y + \omega_3^3\Pi_z] \\
& + [e_0e_1][\omega^0\Pi_x - \omega^1\Pi\ ] + [e_0e_2][\omega^0\Pi_y - \omega^2\Pi\ ] + [e_0e_3][\omega^0\Pi_z - \omega^3\Pi\ ] \\
& + [e_2e_3][\omega^2\Pi_z - \omega^3\Pi_y] + [e_3e_1][\omega^3\Pi_x - \omega^1\Pi_z] + [e_1e_2][\omega^1\Pi_y - \omega^2\Pi_x].
\end{aligned}
$$

The required equations now follow immediately. |

[45]

THE AFFINE CONNECTION OF SPACE-TIME AND CLASSICAL MECHANICS

15. Up to this point we have dealt with the usual notion of equality of space-time vectors. However, the formulae obtained above would continue to be valid for an arbitrary definition of equality of two space-time vectors *whose origins are infinitesimally close*: if a more general definition is used, equations (7) preserve their form *but with modified coefficients*.[13]

Let us assume, as is indeed the case in applications, that the only volume force present is the one due to gravity. If the definition of equivalence of two nearby Galilean frames—or, equivalently, the definition of equality of two 4-vectors whose origins are infinitesimally close—is so chosen as to cancel the gravitational forces, the equations of dynamics would reduce to $G' = 0$. Fix a Galilean frame and choose for e_0, e_1, e_2 and e_3 vectors which remain equal, in the usual sense, to the unit vectors of this frame. Set

$$\omega_0^1 = -X\,dt, \qquad \omega_0^2 = -Y\,dt, \qquad \omega_0^3 = -Z\,dt,$$

and

$$\omega_i^j = 0 \qquad (i, j = 1, 2, 3).$$

Then, the equations of mechanics become

$$\Pi' = 0, \quad \Pi'_x - X[dt\Pi] = 0, \quad \Pi'_y - Y[dt\Pi] = 0, \quad \Pi'_z - Z[dt\Pi] = 0$$

or, equivalently,

13 Note that, if we had attached a *different* Galilean frame, say,

$$\bar{e}_0 = e_0 + ue_1, \ \bar{e}_1 = e_1, \qquad \bar{e}_2 = e_2, \qquad \bar{e}_3 = e_3,$$

at each world point, the formula (7) would have to be modified. In particular, ω_0^1 would have to be replaced by $\bar{\omega}_0^1 = \omega_0^1 + u\omega_1^1 + du$.

$$\Pi' = 0,$$
$$\Pi'_x = \rho X[dtdxdydz],$$
$$\Pi'_y = \rho Y[dtdxdydz],$$
$$\Pi'_z = \rho Z[dtdxdydz].$$

These are the equations of classical dynamics of a continuous medium subject to a volume force which is proportional to the mass. To geometrize gravity, it suffices to choose X, Y, Z to be the components of the acceleration due to gravity.

The result just obtained is completely analogous to the one which led directly to the dynamics of a point particle. Indeed, the formulas

$$d\mathrm{e}_0 = -Xdt\mathrm{e}_1 - Ydt\mathrm{e}_2 - Zdt\mathrm{e}_3, \qquad d\mathrm{e}_1 = d\mathrm{e}_2 = d\mathrm{e}_3 = 0$$

imply that two Galilean frames originating at t, x, y, z and $t + dt, | x + dx, \quad y + dy,$ [46] $z + dz$ should be considered as equivalent if the corresponding triads T and T' are equivalent in the usual sense and T' undergoes a rectilinear and uniform translation of velocity (Xdt, Ydt, Zdt) w.r.t. T.

16. Let us adopt the convention that a given definition of the equivalence of Galilean frames with infinitesimally close origins gives rise to a space-time *affine connection*. It is now easy to see that the gravitational phenomena are compatible with several distinct affine connections on space-time. It is important to note first that, *although the affine connection depends on the matter distribution in space, it does not undergo a substantial change on introduction of a small mass in a given region of space-time*. If the entire system consists only of a small mass, the corresponding affine connection will not depend upon the state of this mass. Any possible modification in the affine connection has the effect that the following terms are added to the expression of G':

$$\left.\begin{aligned} &[\mathrm{me}_1][\bar{\omega}_0^1\Pi + \bar{\omega}_1^1\Pi_x + \bar{\omega}_2^1\Pi_y + \bar{\omega}_3^1\Pi_z] \\ +\,&[\mathrm{me}_2][\bar{\omega}_0^2\Pi + \bar{\omega}_1^2\Pi_x + \bar{\omega}_2^2\Pi_y + \bar{\omega}_3^2\Pi_z] \\ +\,&[\mathrm{me}_3][\bar{\omega}_0^3\Pi + \bar{\omega}_1^3\Pi_x + \bar{\omega}_2^3\Pi_y + \bar{\omega}_3^3\Pi_z], \end{aligned}\right\} \tag{9}$$

where $\bar{\omega}_0^i, \bar{\omega}_i^j$ are the changes in the components ω_0^i, ω_i^j of the affine connection. Thus, the only possible modifications are the ones which make the three terms in parentheses vanish *irrespective of the numerical values of the quantities which characterize the state of the material medium*.

Let us first consider the most general situation in which the affine connection permits two Galilean frames, one with an orthogonal triad T and the other with a *non-orthogonal triad* T', to be equivalent. Let us assume—and it is permissible—that the

vectors e_0, e_1, e_2, e_3 used in the equation (9) are still equal in the *usual sense,* i.e., when everything is referred back to a fixed Galilean frame. Set

$$\bar{\omega}_0^i = \gamma_{00}^i dt + \gamma_{01}^i dx + \gamma_{02}^i dy + \gamma_{03}^i dz,$$
$$\bar{\omega}_i^j = \gamma_{i0}^j dt + \gamma_{i1}^j dx + \gamma_{i2}^j dy + \gamma_{i3}^j dz.$$

The coefficients of the forms Π, Π_x, Π_y, Π_z are

$$\rho, \qquad \rho u, \qquad \rho v, \qquad \rho w,$$
$$\rho u^2 + p_{xx}, \qquad \rho uv + p_{xy}, \qquad \rho uw + p_{xz}, \ldots, \qquad \rho w^2 + p_{zz}.$$

[47] In order to cancel the three terms in parentheses in equation (7), | one can treat them as being independent and focus on one term at a time setting others equal to zero. This yields

$$\gamma_{00}^i = 0, \qquad \gamma_{0j}^i = \gamma_{j0}^i, \qquad \gamma_{ij}^k = \gamma_{ji}^k. \tag{10}$$

These equalities simply express the fact that the three quadratic differential forms

$$\bar{\omega}_0^i dt + \bar{\omega}_1^i dx + \bar{\omega}_2^i dy + \bar{\omega}_3^i dz \qquad (i = 1, 2, 3)$$

vanish identically. One can get the same result from the dynamics of a point particle: the equality

$$\frac{d}{dt}\left(e_0 + e_1 \frac{dx}{dt} + e_2 \frac{dy}{dt} + e_3 \frac{dz}{dt}\right) = 0,$$

as well as equations (7) continue to hold provided one adds to the coefficients ω_0^i, ω_i^j the terms $\bar{\omega}_0^i, \bar{\omega}_i^j$ satisfying:

$$\bar{\omega}_0^i + \bar{\omega}_1^i \frac{dx}{dt} + \bar{\omega}_2^i \frac{dy}{dt} + \bar{\omega}_3^i \frac{dz}{dt} = 0,$$

for all values of the ratios of dx, dy, dz, dt. On the other hand, the results would be different if the components of pressure failed to be symmetric, as is the case when the material is subject to a torque.[14] In this case, expression (9) has to vanish even though

$$p_{xy} \neq p_{yx}, \qquad p_{yz} \neq p_{zy}, \qquad p_{zx} \neq p_{xz}.$$

which implies, as is seen easily, that all the coefficients γ_{ij}^k must vanish leaving only 9 undetermined coefficients instead of 18. In this case, and this case only, does the dynamics of continuous media impose conditions on the affine connection of space-time which are stronger than those imposed by the dynamics of a point particle.

14 This occurs for a magnet placed in a magnetic field.

17. Let us now suppose that the affine connection preserves the spatial metric, i.e., that a reference frame with an orthonormal triad cannot be equivalent to one with a non-orthonormal triad. In this case, we can restrict our field of Galilean frames such that the spatial vectors $\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3$ are everywhere orthonormal. Then the relations (7) continue to hold but with additional restrictions

$$\omega_i^i = 0, \qquad \omega_i^j + \omega_j^i = 0,$$

| which are imposed by the conditions [48]

$$(\mathrm{e}_i)^2 = 1, \qquad \mathrm{e}_i \mathrm{e}_j = 0 \qquad (i \neq j),\ i, j = 1, 2, 3.$$

Furthermore, the three quantities $\omega_2^3 = -\omega_3^2$, $\omega_3^1 = -\omega_1^3$ and $\omega_1^2 = -\omega_2^1$ are now simply the components of the rotation necessary to make the triad T equivalent to T'. Thus, the permissible modifications of the affine connection are dictated by the same conditions as in section 16. However, since now

$$\bar{\omega}_i^i = \bar{\omega}_j^i + \bar{\omega}_i^j = 0,$$

the number of arbitrary coefficients is reduced to four: we have:[15]

$$\bar{\omega}_0^1 = r dy - q dz, \qquad \bar{\omega}_0^2 = p dz - r dx, \qquad \bar{\omega}_0^3 = q dx - p dy,$$
$$\bar{\omega}_2^3 = -\bar{\omega}_3^2 = p dt + h dx, \qquad \bar{\omega}_3^1 = -\bar{\omega}_1^3 = q dt + h dy, \qquad \bar{\omega}_1^2 = -\bar{\omega}_2^1 = r dt + h dz.$$

Furthermore, had the pressure not been assumed to be symmetric, the coefficient h would have vanished.

18. We shall see later on how, following the viewpoint adopted in section 16 or 17 above, one can select among all affine connections compatible with experiments, a specific one, which can be distinguished from others by its intrinsic properties. However, one may consider a theory in which the angular momentum of a typical element of matter about a point located within the element is not negligible compared to its linear momentum, or in which the stress within the medium manifests itself not only via *forces* but also through *torques*. Under these conditions, the analytic expression of G must contain terms such as $[\mathrm{e}_0 \mathrm{e}_i]$ and $[\mathrm{e}_i \mathrm{e}_j]$ and hence *the precise affine connection of space-time would be determined only through experiments;* the experimental data from mechanics would be compatible with only one definition of the equivalence of two Galilean frames whose origins are infinitesimally close.

15 The geometric interpretation of these relations is straightforward.

SPACE-TIME OF SPECIAL RELATIVITY AND ITS AFFINE CONNECTION

19. The theory of special relativity admits the same Galilean frames of reference as classical mechanics. The essential difference lies in the transformation laws between
[49] coordinates (t, x, y, z) which label an | event in one Galilean frame and coordinates (t', x', y', z') which label it in another such frame. These laws are still linear, i.e., the special relativistic space-time continues to be an affine space. However, the time component t of a space-time vector is no longer an invariant. Instead, the invariant quantity now is:

$$c^2(t'-t)^2-(x'-x)^2-(y'-y)^2-(z'-z)^2,$$

where c is the velocity of light in vacuum. As a consequence, the *scalar product*, $c^2\theta\theta'-\xi\xi'-\eta\eta'-\zeta\zeta'$, of two vectors with components $(\theta, \xi, \eta, \zeta)$ and $(\theta', \xi', \eta', \zeta')$, respectively, is also an invariant. In particular, if e_0, e_1, e_2, e_3 denote, as above, unit vectors attached to a Galilean reference frame, the following relations hold:

$$\left.\begin{aligned} &(e_0)^2 = c^2, \qquad (e_1)^2 = (e_2)^2 = (e_3)^2 = -1, \\ &e_0 e_i = 0, \qquad e_i e_j = 0 \qquad (i \neq j = 1, 2, 3). \end{aligned}\right\} \tag{11}$$

20. Consider a variable Galilean frame which depends on one or more parameters. For any infinitesimal variation of these parameters, one has:

$$\left.\begin{aligned} de_0 &= \omega_0^0 e_0 + \omega_0^1 e_1 + \omega_0^2 e_2 + \omega_0^3 e_3, \\ de_1 &= \omega_1^0 e_0 + \omega_1^1 e_1 + \omega_1^2 e_2 + \omega_1^3 e_3, \\ de_2 &= \omega_2^0 e_0 + \omega_2^1 e_1 + \omega_2^2 e_2 + \omega_2^3 e_3, \\ de_3 &= \omega_3^0 e_0 + \omega_3^1 e_1 + \omega_3^2 e_2 + \omega_3^3 e_3, \end{aligned}\right\} \tag{12}$$

where the ω_i^j are linear in the differentials of the parameters and are constrained due to equations (11). On differentiating (11) one easily obtains:

$$\omega_0^0 = 0, \qquad \omega_0^i = c^2\omega_i^0, \qquad \omega_i^j + \omega_j^i = 0 \qquad (i, j = 1, 2, 3). \tag{13}$$

Thus, we are left with six independent quantities, which is precisely the number of parameters required to fix the *orientation* of a Galilean frame.

In the above equations, the coefficients ω_0^1, ω_0^2 and ω_0^3 represent, after change of sign, the (infinitesimally small) uniform, translational velocity of the axes of the second Galilean frame w.r.t. those of the first. Note that, in the limit as c tends to infinity, equations (13) reduce to:

$$\omega_0^0 = 0, \qquad \omega_i^0 = 0, \qquad \omega_i^j + \omega_j^i = 0,$$

| so that one recovers the law relating two infinitesimally close Galilean frames[16] in classical mechanics. [50]

21. *The particle dynamics*. The notion of energy-momentum vector continues to underlie the dynamics of a point particle in special relativity. This vector is now given by

$$m\left(\mathrm{e}_0 + \mathrm{e}_1\frac{dx}{dt} + \mathrm{e}_2\frac{dy}{dt} + \mathrm{e}_3\frac{dz}{dt}\right).$$

The *rest mass* μ of the particle is, up to a multiplicative constant, the square root of the scalar product of the energy-momentum vector with itself. More precisely, we have:

$$\mu = m\sqrt{1 - \frac{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}{c^2}} = m\sqrt{1 - \frac{v^2}{c^2}}.$$

This μ is a number attached to each point particle like the usual mass in classical mechanics.

The mathematical expression of the energy-momentum vector becomes more symmetric if one introduces the *proper time* τ of the point particle, given by

$$d\tau = \sqrt{dt^2 - \frac{dx^2 + dy^2 + dz^2}{c^2}} = \sqrt{1 - \frac{v^2}{c^2}}dt = \frac{\mu}{m}dt.$$

For, the energy-momentum vector can now be written as

$$\mu\left(\frac{dt}{d\tau}\mathrm{e}_0 + \frac{dx}{d\tau}\mathrm{e}_1 + \frac{dy}{d\tau}\mathrm{e}_2 + \frac{dz}{d\tau}\mathrm{e}_3\right),$$

where μ and $d\tau$ are independent of the reference frame.

The fundamental principle of mechanics can be now stated as follows: The derivative of the "energy-momentum space-time vector" w.r.t. the proper time equals the "hyperforce" space-time vector

$$R\mathrm{e}_0 + X\mathrm{e}_1 + Y\mathrm{e}_2 + Z\mathrm{e}_3.$$

The hyperforce vector has an intrinsic significance, independent of the choice of a reference frame: we have |

16 Here, as in section no. 17, the Galilean frames have orthogonal triads.

$$\frac{dm}{dt} = \frac{dm}{d\tau}\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}R,$$

$$\frac{d}{dt}\left(m\frac{dx}{dt}\right) = \frac{d}{d\tau}\left(m\frac{dx}{dt}\right)\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}X,$$

[51]

$$\frac{d}{dt}\left(m\frac{dy}{dt}\right) = \frac{d}{d\tau}\left(m\frac{dy}{dt}\right)\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}Y,$$

$$\frac{d}{dt}\left(m\frac{dz}{dt}\right) = \frac{d}{d\tau}\left(m\frac{dz}{dt}\right)\frac{d\tau}{dt} = \sqrt{1-\frac{v^2}{c^2}}Z.$$

Thus, the force, in the usual sense of the term, is given by the spatial component of the hyperforce times $\sqrt{1-\frac{v^2}{c^2}}$. On the other hand, the constancy of the rest mass introduces constraints among R, X, Y and Z: since

$$c^2 m\frac{dm}{dt} - m\frac{dx}{dt}\frac{d}{dt}\left(m\frac{dx}{dt}\right) - m\frac{dy}{dt}\frac{d}{dt}\left(m\frac{dy}{dt}\right) - m\frac{dz}{dt}\frac{d}{dt}\left(m\frac{dz}{dt}\right) = 0,$$

we have

$$c^2 R dt = X dx + Y dy + Z dz.$$

This relation expresses the fact that the infinitesimal work done by the force equals the change in the quantity mc^2. The quantity

$$mc^2 = \frac{\mu c^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

is the *energy* of the point particle. Indeed, if V is small compared to c, in the first approximation, mc^2 equals

$$\mu c^2 + \frac{1}{2}\mu v^2,$$

or, equivalently,

$$\mu c^2 + \frac{1}{2} m v^2.$$

22. *The dynamics of continuous media*. If we restrict ourselves to the special case
[52] in which all volume forces are absent, the equations | of dynamics of continuous media are essentially the same as in classical dynamics. Thus, we can introduce the sliding vector representing the energy-momentum of an element of matter,

$$\mathrm{G} = [\mathrm{me}_0]\Pi + [\mathrm{me}_1]\Pi_x + [\mathrm{me}_2]\Pi_y + [\mathrm{me}_3]\Pi_z,$$

and set its exterior derivative, G', to be identically zero. If the medium is equipped with a fixed Galilean frame, the components Π, Π_x, Π_y, Π_z can be again expressed as:

$$\begin{aligned}
\Pi &= \rho dxdydz - \rho u dydzdt - \rho v dzdxdt - \rho w dxdydt,\\
\Pi_x &= u\Pi - p_{xx}dydzdt - p_{xy}dzdxdt - p_{xz}dxdydt,\\
\Pi_y &= v\Pi - p_{yx}dydzdt - p_{yy}dzdxdt - p_{yz}dxdydt,\\
\Pi_z &= w\Pi - p_{zx}dydzdt - p_{zy}dzdxdt - p_{zz}dxdydt.
\end{aligned}$$

The density ρ_0 of the matter in its rest frame is given by:

$$\rho_0[dtdxdydz] = [dt\Pi] - \frac{1}{c^2}[dx\Pi_x] - \frac{1}{c^2}[dy\Pi_y] - \frac{1}{c^2}[dz\Pi_z],$$

where the right-hand side is essentially the scalar product of the vector (dt, dx, dy, dz) with the vector $(\Pi, \Pi_x, \Pi_y, \Pi_z)$. On simplifying, one obtains[17]

$$\rho_0 = \rho\left(1 - \frac{u^2 + v^2 + w^2}{c^2}\right) - \frac{1}{c^2}(p_{xx} + p_{yy} + p_{zz}).$$

Thus, in absence of external forces, the equations of dynamics of continuous media are identical to those of classical mechanics. Indeed, if a variable Galilean frame is attached to each point of space-time, one would obtain

$$\begin{aligned}
G' = \quad & [me_0][\Pi' + \omega_1^0\Pi_x + \omega_2^0\Pi_y + \omega_3^0\Pi_z]\\
& + [me_1][\Pi'_x + \omega_0^1\Pi + \omega_2^1\Pi_y + \omega_3^1\Pi_z]\\
& + [me_2][\Pi'_y + \omega_0^2\Pi + \omega_1^2\Pi_x + \omega_3^2\Pi_z]\\
& + [me_3][\Pi'_z + \omega_0^3\Pi + \omega_1^3\Pi_x + \omega_2^3\Pi_y]\\
& + [e_0e_1][\omega^0\Pi_x - \omega^1\Pi] + [e_0e_2][\omega^0\Pi_y - \omega^2\Pi] + [e_0e_3][\omega^0\Pi_z - \omega^3\Pi]\\
& + [e_2e_3][\omega^2\Pi_z - \omega^3\Pi_y] + [e_3e_1][\omega^3\Pi_x - \omega^1\Pi_z] + [e_1e_2][\omega^1\Pi_y - \omega^2\Pi_x].
\end{aligned}$$

23. Let us investigate whether or not several distinct affine connections can be compatible with experiments. In the passage from one | connection to another, the [53]
components ω_0^i and ω_j^i undergo variations, $\bar{\omega}_0^i$ and $\bar{\omega}_0^j$, satisfying, of course, the relations $\bar{\omega}_0^i = c^2\bar{\omega}_i^0$, and $\bar{\omega}_i^j + \bar{\omega}_j^i = 0$. Furthermore, these variations should be such that the four terms

17 Note that, since the volume element $[dtdxdydz]$ is independent of the choice of the reference frame, the quantity ρ_0 has an absolute, frame-independent significance. It is of course not so for the apparent density ρ.

$$[\bar{\omega}_1^0\Pi_x] + [\bar{\omega}_2^0\Pi_y] + [\bar{\omega}_3^0\Pi_z],$$
$$[\bar{\omega}_0^1\Pi\ ] + [\bar{\omega}_2^1\Pi_y] + [\bar{\omega}_3^1\Pi_z],$$
$$[\bar{\omega}_0^2\Pi\ ] + [\bar{\omega}_1^2\Pi_x] + [\bar{\omega}_3^2\Pi_z],$$
$$[\bar{\omega}_0^3\Pi\ ] + [\bar{\omega}_1^3\Pi_x] + [\bar{\omega}_2^3\Pi_y].$$

must vanish identically, *irrespective of the state of the element of matter under consideration*. If the material medium is described using a fixed Galilean frame, one finds, as in section 16, that the four quadratic forms

$$\bar{\omega}_0^i dt + \bar{\omega}_1^i dx + \bar{\omega}_2^i dy + \bar{\omega}_3^i dz = 0 \qquad (i = 0, 1, 2, 3)$$

must vanish identically. One obtains the same expressions for $\bar{\omega}_0^i$ and $\bar{\omega}_j^i$ in terms of four arbitrary coefficients $p, q, r,$ and h as in section 17.

Finally, had we enlarged our notion of mechanics of continuous media by allowing terms of the form $[\mathrm{e}_0\mathrm{e}_i]$ and $[\mathrm{e}_i\mathrm{e}_j]$ in the expression of the energy-momentum density, there would have remained no arbitrary coefficients: *the affine connection of space-time would have been completely determined experimentally.*[18]

24. *Gravitation in special relativity.* In classical mechanics, the equations

$$\omega_0^1 = -X dt, \qquad \omega_0^2 = -Y dt, \qquad \omega_0^3 = -Z dt, \qquad \omega_i^j = 0 \quad (i, j = 1, 2, 3)$$

defining the affine connection which enables geometrization of gravity, preserve their form under the change of the Galilean frame of reference. In special relativity, one may postulate that Newton's law of gravity holds in the Galilean frame whose axes point towards the fixed stars and have the centre of mass of the solar system as origin. However, the law would not have the same form in other Galilean frames. On the other hand, we may follow Einstein and postulate that the law of gravity should have an invariant expression irrespective of the Galilean frame which is used.[19] We are then forced to modify the law itself. Nevertheless, let us note here that the resulting
[54] geometrical | formulation of gravity due to Einstein is essentially similar to the one mentioned in the beginning of this chapter.

25. *The viewpoint of general relativity.* Up to now, we have worked under the assumption that there exist Galilean frames which can label points in the *entire* space-time. At this point, however, it is clear how one can get rid of this assumption.

---

18 This is so if we simply allow torques to act on the elements of matter. For, in that case, the coefficient h is necessarily zero, and, since p, q, r, h transform into one another as components of a 4-vector under Galilean transformations, *one is forced to conclude that* p, q, r *vanish*.

19 The precise meaning of this phrase will become clear later on.

Indeed, to formulate physical laws, it is sufficient that the following two conditions are satisfied:

1) To measure quantities of physical interest, one has available a local reference frame which plays the role of a true Galilean frame[20] in a patch of space-time immediately surrounding the observer, and

2) One knows the space-time connection, i.e., one knows how to compare the observations carried out in two Galilean frames whose origins are infinitesimally close. One may reformulate this condition by saying that one has to know the Lorentz-Minkowski transformation required to make two frames coincide. Analytically, this means that one should know the coefficients in equations (8) and (12).

We shall now go on to the theory of manifolds with an affine connection. Application of this theory to general relativity will follow. We shall also examine the way in which the laws of electromagnetism serve to determine the affine connection of space-time.

20 This is obviously not the place to enter into a discussion of practical difficulties which may arise in assimilating a given reference system to a Galilean frame.

INDEX

Explanatory entries are marked in boldface

A
aberration, stellar 97, 175, 253
Abraham, Max 12, 213, 667, 756, 814, 966
  astronomical consequences of a relativistic theory of gravitation 323–325
  contact with Schwarzschild 165
  controversy with Einstein 305, 422, 545, 609
  discussion of theories of gravitation 363–406
  four-dimensional formalism 193, 236, 287, 492, 925
  objection to Maxwellian theory of gravitation 235, 348
  on Poincaré 214
  theory of electron 221, 254, 790–792, 814
  variable speed of light 12
Abraham's theory of gravitation 311–328, 331–362, 396, 398, 495–496, 517, 665, 667, 677, 730–731, 867
  equations of motion 322
  fundamental equation 356, 376
  Lagrangian 351–353, 378
  see also Mie's theory, Abraham's theory and
absolute differential calculus 470, 557, 612, 964, 1043, 1045, 1058, 1081
accelerated frame of reference, *see* frame of reference, accelerated
accelerating force, *see* force, accelerating
acceleration 376–377, 588–589, 600, 1053
  absolute and relative 611
  field 313
  four-dimensional 290, 490
  gravitational, *see* gravitational acceleration
  relativity of 589
  resistance to 592
action
  equality to reaction 30, 50, 293, 301, 676
  local 1089
  mass 277
  principle of least 254, 280
  retarded, *see* retarded action
  stress 280
action at a distance 1–3, 7, 40, 135, 194, 347, 365, 588, 613–614, 1026
aether 32, 40, 55, 113, 186, 253, 293–294, 613–619, 625, 639, 791, 864, 866, 1056, 1100
  as carrier of electric charge 634
  as carrier of inertia 133, 617
  as foundation of all of physics 634
  currents 102
  density 639
  energy density of 639, 749
  gravitation and 4, 255, 318
  impact 4, 105–110
  in Mie's theory 685, 746
  Lorentz's 22, 55, 194, 591, 788–789
  Mach's 54–55
  of general relativity 55, 57, 617–619
  pressure 639
  represented by metric 1103
  special relativity and 175, 288, 591, 615–617
  stationary 788
  vibrations 103
  vortex 746, 749
affine connection 5, 1041, 1090, 1094–1097, 1103–1105, 1121–1129
  components 1074
  general linear 1073
  in special relativity 1127–1128
  Newtonian 1047
  non-flat 1050
  potential 1054

preserving metric 1123
symmetric 1047, 1075, 1095
uniqueness 1121–1123, 1127–1128
affine group 1062
algebraic
forms 764
number fields 760, 764, 767, 769
α-rays 376
alternate history 1060
analysis situs 1090–1094
analytical mechanics 581, 786, 1042
angular momentum in absolute and relative space 583–585
antinomies of free will 938
Archimedean axiom 770, 796
area element 1092
area law 504
arithmetic, foundations of 773
asteroids 586
astronomy 6, 255, 262, 270, 284, 292–293, 805
general theory of relativity and 52, 165, 910, 954
Newtonian 804, 806
asymptotic flatness 170, 173, 963
atom 590, 866, 901, 933
Bohr 328, 939
interior of 1000
nucleus 327
of electricity 791
structure of 634
atomic configurations 597
atomic physics 900
atomism 596, 776, 806, 820, 832, 846, 849
axiomatic method 760, 776, 781, **840**, 893, 900–901, 921, 959, 989, 1000, 1003, 1015
axiomatization
completeness 770
of arithmetic 780
of *Entwurf* theory 860, 872
of geometry 763, 765, 780, 794
of individual physical theories 777, 784
of mechanics 765, 804, 864
of physics 766, 775, 777–778, 780, 785, 793, 816, 818, 822, 859–860, 864–865, 885, 967, 976–977
of pseudogeometry 940, 1019
of vector addition 815

B

Bacon, Sir Francis 64
balance
ordinary 82, 85
torsion 80–81, 85, 133, 137, 370
basis, orthonormal 1062
Becquerel rays 790
Becquerel, Jean 963
bending of light, *see* gravitational field, deflection of light in
Berkeley, George 573
Bertrand principle 802
Besso, Michele 47, 59, 312, 320, 423
Bianchi identities 859, 862, 893–895, 899, 924, 926–927, 986
Bianchi, Luigi 1045
Birkhoff, Garrett 864
Birkhoff's theorem 949
bivector **1071**
black-body radiation, *see* heat radiation
Blaschke, Wilhelm 1020
Blumenthal, Otto 760, 764, 957
Bohlmann, Georg 776
Bois-Reymond, Emil du 771–772
Boltzmann distribution 824
Boltzmann equation 812, 823
Boltzmann, Ludwig 786, 804, 816
definition of entropy 809–810, 824
foundations of mechanics 803–806, 841
foundations of statistical mechanics 815
kinetic theory 809–812, 823–826
on absolute space 145
Vorlesungen Über die Principien der Mechanik 776, 803–804, 821
Born, Max 287, 292, 295, 788, 819
formalism of relativity theory 243
hyperbolic motion 228
publication of Minkowski's papers 224
reformulation of Mie's theory 623, 631, 745–756, 865–866, 878, 990, 1004
Born-Infeld theory 628
Bosworth, Anne Lucy 779
boundary conditions 51–52, 949
see also Cauchy boundary value condition
Broggi, Hugo 214
Brownian motion 598, 827

Bucherer, Alfred Heinrich 254, 291, 808
bucket model, *see* Newton's bucket
Budde, Emil 129, 144, 298
bundle
  frame, cross-section of 1064
  morphisms 1041
  of linear frames 1063
  tangent 1065
  trivial fibre 1064
Burali-Forti, Cesare 242

C

Cantor, Georg
  continuum hypothesis 773
  set theory 771, 774
Carathéodory, Constantin 819
Cartan, Elie 6, 1047, 1107–1129
Cartesian product 1064
cathode ray 788
Cauchy boundary-value condition 871, 938
  in general relativity 862, 941, 943, 953, 956, 1022
  see also initial-value problem
Cauchy method of residues 324
Cauchy normal form 943, 954
Cauchy, Augustin Louis
  theory of differential equations 887, 991
causality 858, 861–862, 887–888, 900, 934–939, 941, 944, 954–955, 962, 1020
  for generally-covariant field equations 886, 942, 961
  principle 635, 937–938, 941–942, 955–956, 995, 1022–1025
cause, fictitious 611
Cayley, Arthur 216, 224
celestial mechanics 194, 196, 200–201, 214
  initial-value problem of 582
  three-body problem of 582
centrifugal force 33–34, 79, 83, 127–128, 130–134, 137, 160–161, 351, 370, 464, 545, 573–574, 608, 610
centrifugal phenomena 136–142
  invertibility of 127, 133–134, 137
  relativity of 129
characteristics, theory of 941
Christoffel symbol 963, 1047, 1083, 1096
Christoffel, Elwin Bruno 49, 976, 1045, 1057
  transformational calculus 1045
chrono-geometry 1041–1042
classical mechanics 27, 273, 370, 396, 576, 583, 588, 592, 814, 821, 952, 964
  Lagrange equation 583
  of continuous medium 1112–1120
  reformulation of 37
  see also Newtonian mechanics
classical physics 2, 31
  worldview of 41
Clausius, Rudolf 5, 123, 298
clock 287, 319, **581**, 587, **590**, 597, 599, 953
  ideal 1049
  inertial 587
  light 939, 941, 1018–1019, 1024
  microscopic theory of 591, 599
  pendulum 581
closed system 415, 719–720, 723
coincidence, spacetime, *see* spacetime, coincidence
complete static system 14, 415, 445
complete stationary system 415, 457–458, 462–465, 483, 524, 723
condenser 299
conduction, electrical 837–838
conductor 300
  moving 298
connection, *see* affine connection
conservation, *see* energy conservation, energy-momentum conservation, momentum conservation
constrained motion, *see* motion, constrained
constraint, on initial data 955–956
continuity condition 259, 277, 279
continuity equation 751, 824
continuum 805, 1090
  four-dimensional 748, 753
  in special relativity 1126–1128
continuum mechanics 785–787, 804–806, 820, 822, 1112, 1114–1120
  special relativistic 925, 966
Conway, Arthur W. 242
coordinate condition 981
  harmonic 963–964
coordinate restriction 558, 869–871, 881, 887–892, 899–901, 910–914, 919, 922, 981
  energy-momentum conservation and 870–871, 888–893, 901, 910, 913, 919

coordinate transformation 950
admissible 554
continuous 1091
infinitesimal 1028
initial data and 943
invariance 1102
true 1021
see also general covariance, Galilei transformation, Lorentz transformation
coordinates 576, 600, 862, 875, 1090–1091
absolute 129, 131, 143
adapted 891, 903, 1050, 1052
admissible 888, 904
Cartesian 596, 1107
Gaussian 942–944, 948, 954–956, 1022–1024, 1029
geodesic 1096
geodesic normal 961
Kretschmann-Komar 944
physical meaning of 937–938, 940–941, 943–944, 968
preferred 861, 911, 956
proper 941
quasi-Cartesian 866
restricted class of 891
Riemannian 944, 1024
rotating 1052
triad of 1109–1111
true spacetime 941, 1021–1022
Copernican system 142, 158, 184
Copernicus, Nicolaus 208, 255–256, 588
Coriolis force 146, 377, 565, 1052
cosmogony 344
cosmological constant 52, 56–57, 59–60
cosmology 56–57, 59, 161–165, 173–174, 602
Coulomb forces within electron, *see* electron, Coulomb forces within
Coulomb's law 30, 40, 45, 347, 364, 426, 543
covariance, general, *see* general covariance
covariant derivative 1073
creativity 965
Cunningham, Ebenezer 242
curvature 1082, 1084–1088, 1098–1100
extrinsic 1043
Gaussian 1043, 1056
of a connection 1056
of a pseudo-metric 1056
of four-dimensional manifold 1045
radius 1044
curve 1050
as distinguished from path 1065
curved surface 1043

D

d'Alembert principle 801–802
d'Alembert, Jean Baptiste le Rond 1042
d'Alembertian operator 200, 216, 418
Dällenbach, Walter 43
Darboux, Jean Gaston 795
de Sitter solution 52, 59
de Sitter, Willem 52, 60, 173, 241, 293
Debye, Peter 836
Dedekind, Richard 760
deformable body 500
Dehn, Max 779
Desargues, Gérard 765
theorem of 765, 769
Descartes, René 36, 572
dialectics 1043
diffeomorphism 1041
differential calculus, *see* absolute differential calculus
dipole moment, absence in gravitational waves 327
directions, fixed system of 129–132
Dirichlet principle 777, 785
displacement, elastic 748
Doppler effect 288, 350
dragging effect 37, 171
dualism of matter and field 68, 966
Duhem, Pierre 65
duration 349, 585–587
dust, pressureless, as source of gravitational field 903
dynamics
axioms of 579
intrinsic 585
kinematics and 35
of single points 800
Dziobek, Otto 231, 584

E
Earth 80, 293
  irregularities in the rotation rate 585
  motion of 253, 292, 300–301
Eddington, Arthur S. 57, 175, 327, 631, 963
  eclipse expedition 51, 167
  unified field theory 986
Ehrenfest, Paul 47, 313, 439, 806, 815, 819, 881
Ehrenfest, Tatyana 815
Einstein tensor 946–947, 968, 978–979, 985–987
Einstein, Albert
  assessment of Nordström's theory 14
  conception of space 594
  constructive and principle theories 598
  cosmological model 59
  criticism of Mie's theory 631
  elevator thought experiment 46
  Entwurf paper 451
  generalization of Maxwell's theory 626
  heuristic strategy 28, 37, 45, 51, 862–863, 968–969
  Hilbert and 905–906, 957, 959
  November tensor 976
  on aether and relativity 613–619
  on gravitation 543–566
  on the relativity problem 605–612
  philosophical perspective 23, 66
  third way to general relativity 28
  Vienna lecture 623–625
Einstein's equations, *see* gravitational field equation, Einstein's
Einstein-Grossmann theory of gravitation, *see Entwurf* theory
elasticity theory 279, 363, 500–501, 504–508, 746, 748–749
electric charge 1048
  density 256, 635, 751
  positive elementary 397
electric current 635, 751
electric field 750–751
electric potential 639, 990
electricity 1029, 1033
  conservation 960
  four-current density of 1025
  theory of 763, 782
electrodynamic potential 894, 989, 997, 1003, 1005, 1026
electrodynamic worldview 25, 27, 30, 64, 196, 261, 625, 634, 746, 790–791, 845, 905–906, 965
electrodynamics 25, 41, 347, 752, 787, 812–814, 931, 964–966, 1003
  as a consequence of gravitation 860, 885, 894, 897–898, 917, 922, 926, 928, 1000, 1005, 1015
  as foundation of physics 789, 791, 860
  as non-mechanical theory 32
  Clausius' fundamental law of 298
  equations of motion in 813
  foundations of 921, 933, 989
  integration with classical mechanics 966
  Lorentz's 29, 69, 194, 236, 255, 261, 275, 366, 615, 789, 792, 809, 814
  Maxwell's 63, 236–237, 245, 296, 426, 543, 590, 598, 788, 791, 864, 898, 904–905, 935, 1048
  non-linear 628, 866
    see also Mie's theory
  of moving bodies 591, 792, 804
  transition from static fields to full dynamics 29, 426
  unification with gravitation, *see* gravitation, electromagnetism and *and* gravitation, unification with electrodynamics
  see also Maxwell, James Clerk, electrodynamics *and* electron theory
electromagnetic field 55, 258, 260–261, 270, 654, 1049
  as source of gravitational field 655, 858
  gravitational mass of 390
  represented by metric vortex 1103
  tensor 626
  transfer of energy and momentum 385, 391
electromagnetic field equations 289, 291, 297, 364, 400, 533, 634–636, 894, 919
  for ponderable bodies 295
  see also Maxwell's equations
electromagnetic mass 197, 439
electromagnetic origin of matter 907
electromagnetic potential 194, 875, 881
  four-vector 750, 875

electromagnetic wave 113, 347, 376, 390
electromagnetism 821, 859, 879–880, 898, 900–901
  gravitation and 3, 7, 363–364, 901, 910, 914, 927–928
  see also gravitation, electromagnetism and
electron 194, 256, 292, 348, 362, 847, 932
  at rest 657, 1023–1025
  deformable 196–197, 254, 789
  density 298
  dynamics 253, 297, 814
  electromagnetic field within 197, 633
  electromagnetic force on 290
  electromagnetic nature of mass 260, 291, 790
  equation of motion 289–291
  existence 633, 661, 866
  Hamilton function for structure 959
  in Mie's theory 652–662, 668, 670–684, 752, 861
  inertial mass 662
  internal structure 626, 633–634, 638, 655, 959
  longitudinal and transverse mass 291
  mass of 531, 533, 790
  moving 254, 657–659
  polarization and magnetization 299
  radiating 347
  rigid-sphere 254
  shape 291, 530
  surface tension 535
electron theory 175–176, 290, 292, 296, 439, 625–626, 787–788, 790–791, 814, 836–839, 845, 864, 939
  Lagrangian 196–197, 351, 790
  Lorentz's 22, 30, 68, 196, 209, 215, 234, 253–261, 626, 745, 788, 790–791, 814, 836, 838, 965
  Mie's theory and 654
  of metal 827
electrostatics 365
electrotechnology 788
elementary particle 402, 618–619
elevator model 47–49
ellipsoid 254, 257, 263, 291
embedding 1046
energetics 25
energy 294, 391, 908–909, 914–919, 925, 933, 1008
  as source of gravitational field 354
  chemical 351
  density 348, 355–356, 494, 505, 513, 518, 642
  density of gravitational field, *see* gravitational field, energy density
  electromagnetic 347, 351, 385, 873, 909, 995, 997, 1011
  equivalence to mass 11, 26, 354
  flux 494, 496, 505–506, 513, 518, 525, 529, 641, 653–655
  in absolute and relative space 583, 585
  in Hilbert's theory 880–881, 883–885, 890, 892–893, 899–900, 913, 916, 980–982, 994–995
  in special relativity 1126
  kinetic 132, 138, 282, 295, 323, 333, 351, 749, 812
  of solar system 586
  potential 323, 333, 351
  rest, *see* rest energy
energy conservation 25, 282, 294, 322, 334, 342, 368, 370, 372, 397, 452, 492, 494–495, 502, 505, 509, 512, 518, 747, 802, 891, 911, 994–995
  in gravitational field 322
  in Hilbert's theory 916
  in Mie's theory 666
  in Nordström's theory 494
  local 640
  preferred coordinate systems and 861
  problem in scalar gravitational theory 14, 629
energy-momentum conservation 366, 558, 885, 910, 913–914, 919, 961, 966, 968, 1058
  coordinate restriction and 870–871, 888–893, 910, 913, 919
  divergence form 891–892, 914, 944, 1025
  gravitational field equation and 895, 957
  in electrodynamics 756
  in general relativity 861, 919, 923, 926–928, 960
  in gravitational field 544, 556, 880

in Hilbert's theory 919, 930
in Mie's theory 754–756
invariance of the action and 753, 960
energy-momentum expression of the gravitational field 432, 457, 869, 889, 892, 944, 954
energy-momentum tensor **220**–221, 244, 524, 753, 885, 903, 917
Hamilton's function and 961
in Mie's theory 754–755
non-existence for gravitational field 31
of elastic body 500
of general relativity 890, 924–925
of Hilbert's theory 961
of matter 61, 414, 868, 880, 905, 926, 939
of Mie's theory 643, 650–652, 670, 872–874, 879–881, 884, 893–894, 896, 899, 913–914, 916–917, 919, 925–926, 998, 1012
of Nordström's theory 525
of the electromagnetic field 31, 525, 880, 904, 912
energy-momentum vector 1113
in special relativity 1125–1127
entropy 810
Entwurf theory 396, 399, 401–402, 405, 625, 631, 699, 871, 964, 966, 976
Abraham on 320
bucket model in 50
comparison with Mie's theory 867
field equations 868, 903
Hilbert on 888, 893, 935
Lagrangian 868–870
matter in 868
principle of equivalence in 625
relativity principle in 630
envelope theorem 1019
Eötvös, Lorand Baron 545
experiment 420, 434
equilibration process 967, 969
equilibrium 260, 638
chemical 827
principle of 834
equiprobability principle 835
equivalence principle 5, 43, 60, 62, 193, 308, 385, 413, 497, 544–545, 611, 624–625, 968, 976, 1041, 1046–1047
Abraham on 397
as heuristic principle 48, 313, 592, 968
in *Entwurf* theory 625
in Nordström's theory 524, 526
Mach's critique of classical mechanics and 44–45
redshift predicted by 155, 167
theory of static gravitational field and 50
see also Mie, Gustav, criticism of equivalence principle
ergodic hypothesis 827
ether, *see* aether
Euclid 1026
Euclidean
geometry, *see* geometry, Euclidean
group 1061
space, *see* space, Euclidean
Euler, Leonhard 785
approach to continuum mechanics 805
equations of hydrodynamics 787, 808
explanation, reductionistic and phenomenological 820
exploration depth 966, 968–969
extensive quantity 627–628, 636, 650, 731
exterior derivative 1115
exterior product 1117

F

Fano, Gino 761
Fermat cycloids 385
Fermat's theorem 771
fibre 1063
field 931
theory 29–30, 38, 49
theory of gravitation, *see* gravitational field theory
FitzGerald, George 253, 789
Fizeau experiment 614
flow of fluids, stationary 524
flywheel 133, 137
Fokker, Adriaan D. 13, 415, 470
Föppl, August 5, 36, 101, 145–152, 200
gyroscope experiments 148
force
accelerating 282–283, 336–337, 490, 492, 500, 654
as defined by Weyl 1093
centrifugal, *see* centrifugal force

Coriolis, *see* Coriolis force
density 202, 221, 431
driving 222, 227, 229, 240
electromagnetic 254, 261
elimination of concept 37
four-vector 204, 222, 290
gravitational, *see* gravitational force
in Abraham's theory 331
in Hilbert's theory 864
in Mie's theory 652–654, 671–676
in Nordström's theory 490
in Poincaré's theory 259
in special relativity 1125–1126
inertial, *see* inertia
Lorentz transformation of 202, 261–262, 264, 439
mechanical 256, 270
Minkowskian 291
molecular 291, 789
Newtonian 290–291, 293–294
non-electromagnetic 254
of cohesion 638
on conductor 300
on volume element of continuous medium 1114
ponderomotive 221, 227, 500, 510, 524
quasielastic 297
two concepts in special relativity 490
velocity-dependent 36–42, 148–151, 261
Foucault pendulum 128, 132, 140, 159, 188
four-dimensional physics 193, 210, 243–244, 331, 335
see also Minkowski formalism
four-dimensional vector algebra 193, 195, 237, 241
four-vector 424, 490
see also force, four-vector *and* velocity, four-vector
frame
bundle 1041, 1055, 1061, 1063
linear 1062–1063
orthonormal 1062
frame of reference 287–288, 588, 1100
accelerated 44, 60, 1046, 1048
distinguished 598, 600
Galilean, *see* inertial frame of reference
geodesic 1104–1105
inertial, *see* inertial frame of reference
Newtonian 579, 586
non-rotating 1049
Frank, Philipp 242, 295
Fraunhofer lines 350, 398
free fall 341–345, 519
uniqueness of 415
velocity 343
Frege, Gottlob 573, 579, 585, 761, 777
logical system 779
Fresnel, Augustin 253
Freundlich, Erwin 156–157
Friedlaender, Benedict 5, 36, 38, 40, 42, 134–144, 574
law of relative inertia 143
Friedlaender, Immanuel 5, 35–36, 38, 40, 42, 127–134, 574
Friedmann, Alexander Alexandrovich 57, 59
Friedmann's solution 57
functions, theory of 781
fundamental units, dependence on gravitational potential 538

G

galaxy
rotation of 172, 188
structure of 187
Galilei transformation 1048
Galilei, Galileo 36, 184
mechanics 333, 342, 589
Galileo's principle **11**, 26–28, 37, 464, 580
incompatibility with relativity principle 12
invalidity in Nordström's theory 520
Gans, Richard 208, 234, 427, 664, 808
gas
dilute 827
ideal, distribution function 823–824
mixture 826
model 2
theory 822, 826–827
see also kinetic theory of gases
thermal processes in 823
gauge invariance 898, 956, 960
of world function 628
gauge theory 601, 1043, 1059
Gauss, Carl Friedrich 48, 763, 783, 800, 804, 843, 862, 945, 947, 1022, 1025, 1043

measurement of the sum of angles 763
principle of minimal constraint 800, 803
Gauss's theorem 380, 504–505, 540
Geiser, Carl Friedrich 48
general covariance 553, 858, 860–861, 878, 995
uniqueness problem for solutions of field equations 937–938
variational principle and 877, 881, 888, 919
see also hole argument
general relativity principle, *see* relativity principle, general
general theory of relativity 569, 589–601, 610–612, 859, 861–862, 1089
Abraham's theory of gravitation and 398
astronomy and 52, 165, 910, 954
classical mechanics and 1111
cosmological aspects 52, 58
Hilbert's theory of gravitation and 858–863, 893
implications for electrodynamics 930
Lagrangian 859, 958
Machian aspects 50–52, 59, 577, 585
mathematical formulation of 1041, 1081
Mie's critique of, *see* Mie, Gustav, critique of general relativity
Mie's electrodynamics and 873, 998
Newtonian limit 1058
philosophical closure 68
relativism and 587
roots in classical physics 29
variational formulation 859, 873, 928, 958, 967
geodesic 162–163, 1057, 1073, 1097
circular 955
quadrilateral, infinitesimal 1045
time-like 953, 1050–1051
geodesic deviation 1053
geodesic equation 48, 963
geodesic law of motion 37, 868, 951
geodesic lines 950–951, 1019, 1030, 1033
geodesic null lines 940, 953, 1020
geometrical
interpretation 1044–1045
transformations 1061
geometry 349, 759–761, 764, 766, 768–769, 772, 778, 946
absolute 769
algebraic 1060
analytic 760
as model for axiomatic analysis 819
as natural science 763
axiomatization of 762, 764–765, 767–768, 781, 783
Cartesian 770
conformal 1101
empirical foundation of 590, 594–595, 843, 945
Euclidean 210, 596, 763, 765, 768–770, 774, 782, 862, 939, 945, 947–948, 950, 952, 954, 1025–1026, 1061
extrinsic 1067
foundations of 760–761, 774–775, 779, 948
four-dimensional 209
Helmholtzian 596
intrinsic 1067
intuitive 759, 782
local approach to 1044
non-Archimedean 782–783
non-Euclidean 6, 29, 37, 47, 161–162, 226, 761, 769, 782–783, 945, 964, 1044
of spacetime 819, 1110
projective 759–760, 762, 765, 768
pseudo- 1020, 1025
pseudo-Euclidean 940, 950, 1019, 1026, 1029, 1033, 1055
purely infinitesimal 1101
reduction of physics to 597
Riemannian, *see* Riemannian geometry
Gerber, Paul 98–99
Gibbs, Josiah Willard 794, 808–809
Grassmann, Hermann 6, 761, 1044–1045, 1079–1080
gravitation
absolute motion and 261
as a field effect 110–111
dependence on distance 91–92
dependence on mass 87–91
dependence on medium 92–93
dependence on time 93–94, 145, 150
electrodynamic explanation of 9, 29–30, 40, 110, 119–126, 198, 789, 839, 848, 866

unification with electrodynamics 56, 58, 789, 859–860, 867, 872, 905–906, 908, 932–933, 939, 957–958
electromagnetism and 7, 900, 910, 914, 920, 927–928, 964
existence of matter and 633, 667, 866
field-theoretic reformulation of Newtonian theory 8
fundamental role in the structure of matter 904
geometrization of 1121
in Mie's theory 663–667
in special relativity 254–255, 282, 292–294, 821, 1128
inertia and 5–6, 134, 159, 499, 967–968, 1041
light and 991, 1005
Lorentz-covariant theory 195, 413, 571
mechanical explanation of 4, 102–110
modified law of 10, 293
Newton's law, *see* Newton's law of gravitation
Newtonian theory 1–4, 24, 293, 322, 413, 426, 613, 805, 809, 861, 870, 901, 966, 968, 1041, 1047
Nordström's theory, *see* Nordström's theory of gravitation
of energy 310
propagation 94–95, 126, 254–255, 261–262, 347, 361–362
propagation velocity 199, 204, 209, 225, 265, 284, 292, 366, 789
relativistic theory of 292, 331–339, 366, 489–497, 515–521, 523–542, 608, 821
thermal theory of 3, 7
gravitational acceleration 79, 520
dependence on horizontal velocity component 389, 542
dependence on rotation of bodies 542
gravitational constant 79–80, 345, 360, 489, 528, 677, 879, 1053
calculation 83–86
in Mie's theory 684
gravitational factor **515**, 523–524
gravitational field 349, 356, 664
affine connection and 1096
as its own source 31, 50
deflection of light in 13, 51, 165, 167, 193, 310, 322–323, 327, 333, 398, 478, 870, 907, 952
energy density 335, 338, 356, 365, 371, 380, 427, 494, 677
energy-momentum, *see* energy-momentum expression of gravitational field
excitation of 628
homogeneous 519–520
in *Entwurf* theory 554–557
in Mie's theory 651, 664, 675–682, 701
lines 681
momentum density 339
Newtonian limit of 560–563
of moving particle 679–682
representation by a vector 426, 664
source of 14
state of the aether in a 700
static 46–48, 50, 519, 524, 976
stationary 45, 335
superposition principle 676
transfer of energy and momentum 334, 385, 391
vanishing in suitable coordinate system 1097
gravitational field equation 401, 559–560, 709, 857–858, 924, 994, 998, 1042, 1053
derivation using Lagrangian 867–868
Einstein's 949, 975–977, 987, 1041
exact solutions of Einstein's 168–169
explicit form of the 922–923
integrability condition 895
interdependence of the equation of motion and 31
gravitational field theory 9–10, 29, 41–42, 44, 234–235, 629, 1048
negative energy problem 10, 729–730
scalar **11**, 547–552, 663, 730
tensor **11**, 399, 663, 730
vector **11**, 234, 363–366, 664, 729
gravitational force 261, 331–339, 347, 364, 371, 377, 379, 387, 392, 394, 403, 405, 515, 672–676
four-dimensional 308
see also tidal force
gravitational induction 47, 675

gravitational potential 331, 342, 489, 515, 680, 989, 1053
four-dimensional 48, 308, 628
in Mie's theory 390, 629
length, dependence on 533–535
mass, dependence on 490–491, 516, 523, 529
role in gravitational theory 10, 729–731
speed of light, dependence on 331, 333, 341, 396–398, 489
time development, dependence on 536
wavelength, dependence on 538
gravitational redshift 155, 166–167, 322, 953
gravitational tensor 335, 337, 369, 391, 401, 495–496
gravitational wave 207, 322, 325, 327–329, 332, 362, 376, 389, 685–694, 696, 947
from radioactive decay 328
linearized 961, 963
shock 950
Grossmann, Marcel 48, 50, 212, 589, 908, 911
group theory 254
guiding field 1047

H

Hall, Edwin Herbert
free fall experiments 149
Hamel, Georg 779, 795, 797, 815, 819–820
Hamilton, William Rowan 1044
Hamilton's principle 279, 295, 644, 667, 803, 813, 935, 945, 962, 990, 1004, 1026
Hamiltonian formalism 627, 864, 963, 1020
Hamilton-Jacobi equation 939–941
Hargreaves, Richard 193
heat
conduction 499, 510–513
conductivity tensor 510–511
flow, *see* rest heat flow
transport 524
heat radiation 787, 828
Heaviside, Oliver 198, 234, 664, 794, 808
Helmholtz, Hermann von 63–65, 186, 590, 594–595, 775, 785, 787, 1046, 1079
rigid bodies 596, 598, 600
Helmholtz-Lie problem 775
Herglotz, Gustav 213, 292, 324, 447, 499, 746, 748–749, 788, 792
Hermite, Charles 212
Hertz, Heinrich 64, 804, 809
electrodynamics 40, 614–615
mechanics 6, 37, 209, 762–769, 776–777, 798, 800, 803–804, 816–817, 821, 841
Hertz, Paul 792, 885–886
Hessenberg, Gerhard 1045
heuristic strategy, *see* Einstein, Albert, heuristic strategy
Hilbert, David 9, 213, 226, 789, 857, 949–969, 975–1038
adoption of Einstein's energy-momentum tensor 928
axiom for light propagation 1037
axiom of continuity 762, 796, 818
axiom of general invariance 875, 899, 990, 997, 1004
axiom of space and time 888, 899, 995
axiomatic approach to physics 773, 863, 871
axiomatic method 759–850
axioms defining the state of equilibrium 797
axioms for mechanics 815
axioms of general relativity 874, 951, 989, 991, 1000, 1003, 1015
basic equations of physics 1017
competition with Einstein in discovery of gravitational field equations 975–979
conception of matter 951
correspondence with Einstein 903
correspondence with Felix Klein 761, 923
deductive structure of the Proofs theory 899–901, 913
deductive structure of theory of First Communication 917, 928, 930
energy condition 891
energy conservation 801, 880–881, 890, 901, 917, 920
energy in unified theory 861, 872, 885, 919–920, 925
fifth problem 775
foundation of physics 857–969
fundamental equations of gravitation 1005
generalized Maxwell equations 980–982, 984, 1005

gravitational action integral 975, 977–979
gravitational and electromagnetic field equations 857, 926
Lagrange equations of unified theory 991, 1004–1005
Lagrangian 874, 876, 885–886, 895, 912, 914, 917, 926, 930, 961
Leitmotiv 899, 901, 921–922, 931
light ray axiom 955
Mie's axiom of the world function 899, 990
on Mie's theory 631, 977
Proofs 780, 858, 860, 872, 874–876, 880, 884–902, 909–913, 917, 920–923, 925, 931, 934, 1001
reception of his theory 964
seminars on mechanics 765
sixth problem 775–778, 825
twenty-three problems 976
Hofmann, Wenzel 36, 569, 574–577, 584–585, 587, 589, 593, 600–601
hole argument 53, 860, 862, **869**–871, 885, 888–889, 910, 922, 963
holonomic base 1075
Hopf, Ludwig 314, 316, 423, 849
Hubble, Edwin Powell 59–60
Hume, David 63–64
Huntington, Edward 768
Hupka, Erich 291
Hurwitz, Adolf 212, 764
Huygens, Christiaan 36, 573, 593
Huygens' principle 323, 333, 349, 398
hydrodynamics 363, 808
hydrogen atom 909
hydrostatic pressure 338
hyperbolic partial differential equations 886
hyperbolic shell 274
hypersurface 1075
rigged 1050

I

induction 29, 41, 300
inertia 137, 143, 145, 159, 186, 405, 572, 576–577, 588–589, 592, 1048, 1052
gravitation and 5–6, 134, 159, 499, 967–968, 1041
law of 129–143, 577, 580, 804, 816, 1107–1109
of energy 369, 389, 499, 608
origin of 5, 160, 254
relativity of 127, 138–141, 405–406, 470, 563–565, 612
inertial frame of reference 580, 587–588, 599, 1047, 1107, 1109–1129
global and local 598, 1128–1129
in special relativity 1124
infinitesimally close 1108
Mach's view on 34
non-uniqueness 1111
preferred 1128
privileged status of 36
rotation and 1111–1112
inertial mass, *see* mass, inertial
inertial motion 143, 580, 1049
generalized 312, 1049
inertial spatiotemporal framework 581
inertial system 6, 145, 588, 592, 596, 599
local approximations to 597
see also inertial frame of reference
inertio-gravitational field 48, 1041–1042, 1049
Newtonian 1054
initial data 600
constrained 956
initial hypersurface 942
initial-value problem 582, 600, 862
of dynamics 583
integral equation 778, 825
linear 785, 825
theory of 824–825
intensive quantity 627, 636, 638, 641–642, 648–649, 731–732, 750
invariance group 1055
invariant 878, 944, 1045
theory 771, 887, 901, 1001, 1057
invariant statement 953–954
irrational numbers, theory of 773
Isenkrahe, Caspar 4
Ishiwara, Jun 243

J

Jacobian 258
Jaumann, Gustav 7
Jeans, James 828, 830
Joule heating 431
Jupiter, eclipses of moons 293–294

K
Kant, Immanuel 63, 128, 144, 767
  kinetics 129
  phoronomics 129
Kaufmann, Walter 214, 788, 790–791
  measurement of electron mass 254, 790
Kepler, Johannes 255, 588
  equation 230
  laws of planetary motion 951
  motion 230–231, 234, 377, 1035
  second law 208
Killian, J. W. 428
kinematics
  dynamics and 35
  relativistic 217
kinetic theory of gases 363, 776, 809–811, 816, 822–824, 833–837
Kirchhoff, Gustav Robert 64, 787
Kirchhoff's law 825, 828, 831
Klein, Felix 157, 236, 921, 930, 986
  geometry 243
  mechanics 212
  on general relativity 860, 892–893, 898, 919, 924, 927, 960–962
  on Hilbert's achievements 859, 880–881
knowledge
  accepted 1042
  epistemic structures of physical 968
  integration of 65, 952, 964–968
  shared 321–322, 964–965, **967**–969
Koch, K. R. 145, 150
Koebe, Paul 214
Kollros, Louis 212
Komar, Arthur 944
Kottler, Felix 206–207
Kretschmann, Erich 53, 320, 396–397, 570, 944
Kronecker, Leopold 211, 772, 774, 781
Kuhn, Thomas 241

L
Lagrange equation 746, 777, 801–803, 841, 886, 1035
Lagrange, Joseph-Louis 582–584, 785, 787, 1042
Lagrangian 194, 352, 626–628, 745, 808, 865, 867–868, 870, 983, 1004
  energy-momentum conservation and 880, 919
  for equation of motion in special relativity 295
  split into gravitational and electrodynamical terms 899
  variational derivative **872**, 881, 885, 894, 897, 996
Lagrangian approach 208, 745, 806, 820, 871
Lagrangian derivative 883, 963, 992, 1008
Laguerre, Edmond 210
Lange, Ludwig 5, 33, 573, 578–589, 596, 599, 800, 1055
  construction of the spatial frame of reference 579
Langevin, Paul 197, 254, 423
Laplace axiom 813
Laplace equation 313, 357, 385, 398, 807
Laplace operator 418, 807
Laplace, Pierre-Simon de 83–84, 97, 100, 126, 196, 208, 254, 265, 292–293, 366, 813–814
Larmor, Joseph 216, 788–789, 798
Laub, Jakob 222, 242
Laue scalar 14, 476, 524
Laue, Max von 27, 156, 168, 213, 224, 500, 588, 597, 599, 756, 788, 955
  four-dimensional formalism 492, 925
  relativistic mechanics 415, 437, 499–500, 503, 524
  textbook on relativity 242–243, 437
Laue's theorem 525, 657–659, 677, 703, 720, 722–723, 726, 735
laws of nature 253, 607
Le Sage, Georges-Louis 4, 105–109, 113, 198, 848
Leibniz, Gottfried Wilhelm 36, 573, 593
Lemaître, Georges 59, 327
length
  contraction 198, 595, 667
  measurement, *see* rod
  transfer of unit of 1101
  unit of 349
Leverrier, Urbain 907
Levi-Civita, Tullio 6, 292, 470, 976, 1045, 1081–1088
  general linear connection 1073

Lewis and Tolman's bent lever 438
Lewis, Gilbert Newton 193, 236, 437, 840
Lichnerowicz, André 948
Lie bracket 1075
Lie derivative 875–876, 884
  of the Lagrangian 899
  of the metric tensor 881
Lie variation 980
Lie, Marius Sophus 761, 775, 876
Liénard, Alfred 194, 300
  force 301
Liénard-Wiechert law 232
light 292, 362
  cone 226, 232, 239, 282, 601
  deflection of, *see* gravitational field, deflection of light in
  emission theory of 204, 323, 333
  pendulum 843
  point 282
  speed of, *see* velocity of light
line element 595
  four-dimensional 311
linear form 1092
linear space 1060, 1091
linear transformation 1062
Liouville theorem 815–816
Lorentz contraction 209, 253, 255, 392
Lorentz force 226
Lorentz group 261, 268, 283, 320, 362
Lorentz invariance 200, 263, 377, 386–387, 506, 512, 598, 643, 749, 789, 821, 836
  of Mie's theory 647–651, 667–668, 752
Lorentz model of a field theory 3, 965
Lorentz scalar 866
Lorentz transformation 254, 256–260, 267, 273–275, 283, 288, 316–317, 814, 821, 844
  as rotation 263, 366, 400
  for heat 432
  in the infinitesimally small 314–315
  physical significance of 195, 1056
Lorentz, Hendrik Antoon 194, 198, 206, 236–237, 313
  aether, *see* aether, Lorentz's
  electrodynamics, *see* electrodynamics, Lorentz's
  electron theory, *see* electron theory, Lorentz's
  on general relativity 893, 958, 960, 963
  on special relativity 222, 233, 287–301
  Rome lecture 830
  St. Louis address 196
  theory of aberration 175
  theory of gravitation 9, 97, 113–126, 198, 234, 292, 364, 427, 664, 848
  Wolfskehl lecture 830, 836

M

MacCullagh, James 746, 749, 866
Mach, Ernst 5, 35, 135, 570, 573, 576, 584, 589–595, 776, 787, 800, 962
  aether, *see* aether, Mach's
  critique of classical mechanics 21, 23, 27–28, 33, 35, 42, 45–46, 54, 60–61, 145, 576, 584
  definition of mass 34, 37, 42, 46
  economy of thought 406
  inertia 22, 44, 50–51, 577, 624
  Newton's bucket experiment 34, 36
    see also Newton's bucket
  relation to Einstein's physics 22, 60, 63, 65, 570, 946
  relativity of space 43, 129, 569, 573, 592, 600, 617
Mach's principle 42, 53–61, 470, 569–602
  Einstein's introduction of 53–54
Machian defect 602
Machian theory of motion 577, 584, 589
Madelung, Erwin 848
magnetic field 750–751
  static 45, 47
magneto-cathode rays 209
manifold 1010, 1020, 1041, 1044, 1056–1057, 1063–1064, 1074, 1076, 1081–1083, 1089–1090, 1094, 1096, 1098–1104, 1115, 1129
  affine **1112**
  affinely connected 1046
  base 1063
  n-dimensional 1090–1091
  spacetime 215, 274, 321, 424, 557, 875, 932, 979, 983, 1010, 1045, 1059, 1112
Marcolongo, Roberto 210, 242
mass 138, 499, 510, 523
  center of 583
  conservation 275–277

density 520, 1053
dependence on temperature 683, 696
distant 610
equivalence to energy 11, 26, 354
gravitational 197, 351, 376–377, 389–390, 393, 415, 499, 526, 530, 534, 675, 1048
gravitational and inertial 26–27, 42–43, 370, 413, 489, 608–609, 676, 682–684, 719–723, 735–742, 1046, 1110
in volume element of continuous medium 1113
inertial 43, 351, 389, 415, 499, 529–530, 532, 570–679, 1048
negative 101, 662
point, *see* point mass
rest, *see* rest mass
variability of 13, 490, 496, 507–508, 517, 529
material point 539, 542
material tensor 391, 401, **501**, 511, 524
mathematical strategy, *see* Einstein, Albert, heuristic strategy
matter
constitution of 832–836, 958–959
electromagnetic theory of 255, 904–905, 907, 976–979
molecular theory of 832–836
Maxwell distribution 810–811, 825
Maxwell stress 220, 279
Maxwell tensor, *see* electromagnetic field, tensor
Maxwell, James Clerk 40, 194, 293, 427, 614–615, 823, 826
electrodynamics, *see* electrodynamics, Maxwell's
Maxwell's equations 209, 215, 224, 234, 242, 597, 635, 745, 747, 750–751, 808, 846, 878, 898, 955, 1011
applied to gravitation 97, 364, 729
as a weak-field limit, in Mie's theory 626, 697
gravitation and, in Hilbert's theory 887, 893, 908, 935, 991, 998, 1012, 1023
Maxwell-Boltzmann collision formula 824
mechanical worldview 62, 791, 818
see also mechanics, as foundation of physics
mechanics 25, 31–32, 759, 766, 786–787, 793, 809, 812–813, 819–820
alternative systems of 840
as foundation of physics 134–135, 820, 846–847, 849–850
difference between classical and relativistic 504
for non-Euclidean geometry 37
foundations of 31, 35, 133, 594, 625, 630, 803, 841, 1055
generally relativistic theory of 28
heretical 3, 5–6
principles of 30, 798, 815, 1044
reversibility of the laws of 810
see also analytical mechanics, classical mechanics, continuum mechanics, Newtonian mechanics, relativistic mechanics, statistical mechanics
mental model **2**
see also elevator model, gas model, Lorentz model, Newton's bucket, umbrella model
Mercury, perihelion anomaly, *see* perihelion anomaly, of Mercury
metric 875, 1084, 1089–1090, 1100–1103
for a static gravitational field 1059
regular 950
see also Minkowski metric
metric tensor 48, 857, 875, 912, 935, 941
measurement 939–940
Michelson aether drift experiment 253, 292, 301, 667
microphysics 862, 933
Mie, Gustav 8, 477, 623–631, 920, 932, 961, 963
criticism of equivalence principle 630–631, 704–705
critique of general relativity 623–624, 630, 727–728
electrodynamic worldview 235, 745, 885, 894
Mie's theory 9, 623–631, 633–756, 819, 861, 864, 866–868, 870, 873, 878, 880, 898, 962, 966, 977, 989–990, 1000, 1003–1004
Abraham's theory and 628, 651, 665, 696
axiom of the world function 874, 1004
Einstein's theory and 696, 699–728, 873, 926, 1012

empirical predictions 26, 696, 865
energy-momentum tensor, *see* energy-momentum tensor, of Mie's theory
existence of electron 745
field equations 645–647
Hamiltonian 642, 645, 649, 651, 701–702, 731
see also Mie's theory, Lagrangian
Hilbert on 877, 882, 885, 900, 921, 935, 977
influence on Hilbert 631, 846–847, 857–860, 871, 893, 901, 959
invalidity of equivalence principle 389, 609, 738
Lagrangian 626–627, 642–644, 865, 867–868, 871–873, 878–880, 893–894, 896, 899–900, 920, 925, 930, 990, 1015
of gravitation 241, 402, 628–629, 663, 665–697, 906
relativity of gravitational potential 390, 630, 694, 696, 714–719, 729, 732–735, 742–743
Weyl on 960–962
Mie-Nordström theory 389
Mill, John Stuart 64
Minkowski formalism 38, 48, 305–307, 311, 319, 334, 400, 492, 925
Abraham's modification of 306, 332
Minkowski metric 316, 946–947, 950
uniqueness 862, 948
Minkowski spacetime 51, 311, 414, 880, 946
in rotating coordinates 49–50
Minkowski, Hermann 193, 287, 295, 500, 756, 761–762, 771, 774, 786–793, 804, 864
Cologne lecture 236, 239
four-dimensional formulation of special relativity 10, 285, 316, 332, 544, 1055
on electrodynamics 490, 819, 825
relativistic law of gravitation 10, 211–235, 273–284, 416, 821
Minkowskian force 222, 290
Minkowskian mass 222
Mittag-Leffler, Gustav 212
momentum 391, 499, 638
density 496, 503–504, 507, 513, 518
electromagnetic 196
total 524
momentum conservation 322, 327, 333, 366, 368, 372, 502
Monge's differential equation 940, 1020
Monge-Hamilton theory of differential equations 1020
Moon, anomalies in the observed motion of 585
Moore, Eliakim Hastings 768
Mossotti, Ottaviano Fabrizio 8, 119
Mossotti's conjecture 198
motion 135, 572, 578, 590, 1093
absolute 127, 135, 145, 253, 261, 569, 573
circular 536
constrained 37
inertial, *see* inertial motion
interior 520, 540, 737–738, 741, 743
of isolated particle 136
phenomenological 128
proper 185, 187
quasi-stationary 540
relative 36, 135, 138, 145, 569, 572
Müller, Conrad 214

N

n-body problem 582
negative energy problem, *see* gravitational field theory
neo-Kantianism 65
Nernst, Walter 787
Neumann, Carl 5, 32, 100–101, 573, 578–579, 581, 584–585, 587, 589, 798, 800, 816, 1055
body alpha 32, 318
inertial clock 578, 599
Neumann-Lange-Tait procedure 581, 586, 596
Newstein, Isaac Albert 1042
Newton, Sir Isaac 32, 36, 128, 194, 569, 572–573, 579, 582, 584, 587–588, 600
Philosophiae naturalis principia mathematica 79
philosophical presuppositions 31, 35
Newton's bucket 31, 42, 44–45, 48–50, 136, 140, 573
Newton's law of gravitation 45, 79, 151, 193–194, 199, 204, 206, 208, 228, 231–232, 235, 239, 255, 262, 266–268, 270,

282, 284, 292, 324–325, 347, 364, 366, 381, 543, 582, 676, 759, 808, 821, 908, 951–952, 1038, 1048, 1111, 1128
analogy to Coulomb's law 29
for infinitely large masses 100–101
for moving bodies 95–99
Lorentz-covariant generalization 26
tests of 86–92
Newton's laws of mechanics 578–579, 582, 585
first law, *see* inertia, law of
second law 308, 577, 800
third law, *see* action, equality to reaction
Newton's theory 593, 946
Newtonian limit 29, 904, 907–908, 910, 912, 946, 952, 966–968, 1053, 1059
Newtonian mechanics 32, 37, 41, 61, 274, 322, 324–325, 570, 573, 575, 579, 583–585, 595, 788, 892, 1107–1112
equation of motion 582–584, 801, 813, 821
relational formulation 582
Noether, Emmy 864, 888, 919, 921, 931
Noether, Fritz 292
Noether's theorem 859, 893, 986
non-Euclidean geometry, *see* geometry, non-Euclidean
non-Euclidean physics 1025–1026
Nordström, Gunnar 12–13, 27, 414, 571, 593, 609
electron model 463, 530
Nordström's theory of gravitation 388, 394, 396, 404, 489–542, 546–552, 731–732
as modification of Abraham's theory 515
comparison with Mie's 628
dependency of physical quantities on gravitational potential 459–463
Einstein's objections 496, 533
equations of motion of a mass point 539–542
first theory 389, 724–725
force 490
gravitational factor 515
gravitational source 526–527
inertial mass, definition 509
Lagrangian 388, 394
second theory 392–393, 725–726
null cone 1020
null line 1018, 1104
number theory 764, 771, 811

O

Ohm's law 216
Olbers' paradox 6
optics
geometrical 803
unification with electrodynamics 791
orthonormal vectors 1051, 1065
oscillation
electromagnetic 844–849
of material particle 537, 685–694

P

Pappus's theorem 765, 769
parallax, annual 162–163
parallel axiom 763, 769, 1026
parallel displacement 1041, 1045, 1047, 1083, 1090, 1094, 1096, 1104–1105
infinitesimal 1091, 1101, 1103–1104
parallelism 1043, 1061, 1082
particle
elementary, *see* elementary particle
force-free 588, 599
in Mie's theory, *see* electron, in Mie's theory
particle-like solutions in field theory 626–628
Pasch, Moritz 760–761
path
as distinguished from curve 1065
principle of the straightest 803
Pauli, Wolfgang 221, 233, 478, 949, 962
on Mie's theory 626–628
Pavanini, G. 325
Peano, Giuseppe 761
pendulum 83
double 81–85
light 843
see also Foucault pendulum
perihelion anomaly 4, 6, 9, 159, 165, 167, 189, 293, 952
of Mercury 30, 100, 125, 209, 293, 324–325, 469, 861, 870, 906–907, 909–910, 966, 976, 1035, 1081
perpetuum mobile 14, 598, 801, 835

philosophy, influence on physics 21, 62
physical constants, reduction to mathematical constants 901, 1000, 1015
physical strategy, *see* Einstein, Albert, heuristic strategy
physically meaningful statement 937–938, 942–943, 956, 1024–1025
physics
  foundations 900, 989
  fundamental equations 989, 995
  geometry and 901, 945, 1000, 1015, 1089
  laws 938, 1024
  unification 8–9, 931
Pirani, Felix 60, 570, 592
Planck constant 829
Planck, Max 64, 167, 213, 222, 236, 241, 437, 809, 816, 828, 830, 836
  discovery of the quantum of action 590, 597
  four-dimensional formalism 925
planetary motion 86–88, 91, 94–95, 99–100, 108, 124–126, 151, 156, 159, 166, 171, 189, 294, 324–325, 389, 416, 476–477, 951, 964
  see also solar system
Poincaré pressure 197, 626
Poincaré stress, *see* Poincaré pressure
Poincaré, Henri 63, 65, 197, 569, 582, 585–586, 589, 601, 792, 814, 830, 947
  on absolute space 145, 573, 578–579, 581–584, 590, 594
  relativistic law of gravitation 10, 193, 253–271, 282, 293, 416
  St. Louis lecture 196, 214
  theory of cycles 1037
  work on relativity 10, 300
point mass 170, 200–201, 204, 224, 281–284, 327, 333, 378, 381, 383, 417–420, 431–432, 435, 439–440, 446, 451, 464, 470, 492, 496, 553, 563, 951, 1033, 1035–1037, 1107–1111, 1113
Poisson equation 8, 49, 417, 807, 1054, 1111
  four-dimensional 307, 310, 518
polarization **875**, 914
Pomey, Jean-Baptiste 200
popular scientific literature 67, 69
postmature concept 1043
Poynting vector 220, 348
Prandtl, Ludwig 808
pressure
  as intensive quantity 638–639
  in continuous medium 1115
  on a moving surface 291, 503
  see also Poincaré pressure
probabilities, calculus of 776, 809–812, 822
projection operation 1063
Ptolemaic system 142, 255
Ptolemy, Claudius 208, 256
Pythagoras's theorem 595

Q

quadratic differential form 595, 1045
quantum gravity 1041
quantum physics 56, 58, 599, 830–831, 923, 932
quaternion 212, 222, 228

R

radiation
  cavity 598
  diffuse 848
  of energy 347
  Planck law of 829
  Rayleigh-Jeans law of 829–830
  theory of 787, 822, 824–825
  Wien law of 828
  see also heat radiation
radioactivity and proportionality of inertial and gravitational mass 351, 370, 390
ray, magneto-cathode 256
Rayleigh, Lord (Strutt, John William) 787, 828
real numbers, proof of the existence of the continuum of 774
reductionism 837–839, 849–850, 864
reference body 578–579
reflection, laws of, and motion of the Earth 253
refraction
  coefficient 802
  double 301
  laws of, and motion of the Earth 253
regularity of a function in physics 950
Reiff, Richard 236, 805, 807
Reissner, Hans 5, 36, 569, 575–577, 584–585, 587, 589, 593, 600–601

relativism, Cartesian 572, 579
relativistic mechanics
  Lagrangian 574–575
  of deformable bodies 499
  of stressed bodies 415
relativity of inertia, *see* inertia, relativity of
relativity of simultaneity, *see* simultaneity, relativity of
relativity principle 291
  as foundation of electrodynamics 273
  general 362, 553, 589–590, 857, 945, 959, 1023, 1026
  generalization of 320, 402, 552–565, 592, 594, 598, 624, 630
  gravitation and 270, 332, 489, 544, 664
  mechanistic generalization of **28**, 31, 35–36, 38, 40–41, 44–45, 49, 53, 62–63, 67
  special 253–255, 279, 282, 287–294, 297–298, 300–301, 377, 438, 570–571, 583–584, 589–591, 596, 601, 605–607, 634–635, 642, 696, 731, 752, 787, 808, 821, 825, 841, 940, 1019, 1055
  theory of gravitation and 667
reparametrization invariance 601, 936, 1024
rest energy 393, 529, 533
  density 506, 509–510, 520, 527
rest heat flow 512
rest mass 222, 290, 295, 1125
  density 220–221, 276, 492, 500, 509–510, 515, 523–524
  inertial 415
rest volume 502
retarded action 194–195
retarded potential 432, 518, 528
Ricci scalar, *see* Riemann curvature scalar
Ricci tensor 904, 976, 978, 987, 1053, 1059
Ricci's lemma 1085
Ricci-Curbastro, Gregorio 976, 1057
  absolute differential calculus 470
Riecke, Eduard 213
Riemann curvature scalar 414, 859, 928, 978–979, 984–987
Riemann tensor 472, 903, 922, 944, 1010, 1053, 1070, 1081
Riemann, Bernhard 49, 595, 775, 862, 945, 976, 991, 1005, 1079, 1082, 1086–1088
Riemannian geometry 959, 1041, 1044
Riemannian parallelism 1046
Riemannian space 311, 601, 1046
rigging 1050, 1072
rigid body 292, 590, 595, 797, 838, 864
  empirical realization of geometry by 595
Ritz, Walter 204, 212
Ritz's experiment 792
rod 287, 319, 595–597, 599, 939, 941, 953, 1018
  ideal 1049
  microscopic theory of 591, 599
rotation 136, 292, 520, 524, 540, 655, 1061
  absolute 132, 1055
  in a gravitational field 497
  in an otherwise empty space 142
  in gravitational field 27
  of element of continuum 749–750
  of the Earth 572, 585
  relativity of 158, 169
  stability of axis 128, 132
Runge, Carl 797, 892–893
Russell, Bertrand 779

S

Sackur, Otto 827
scalar product 201, 203, 1100
scale invariance 601, 960
Schimmack, Rudolf 795
Schlömilch, Oscar Xavier 797
Schmidt, Erhard 214
Schouten, Jan 876, 1045
Schrödinger, Erwin 5, 36, 575–577, 584–585, 593, 601
Schur, Friedrich 763, 795, 797
Schwartz, Hermann M. 201
Schwarzschild metric 155, 862, 939, 946–951, 953, 955, 961, 963
  variational derivation 959, 1030–1032
Schwarzschild radius 155, 326, 955
Schwarzschild singularity 327
Schwarzschild, Karl 7, 155, 157, 183–189, 194, 213, 233, 323, 792, 814, 860, 949, 952, 955, 1029, 1033, 1035
Seeliger, Hugo von 6, 100–101, 157, 160, 162, 208, 293
Silberstein, Ludwik 243
simultaneity 288, 586–587, 590
  relativity of 442

singularity 325, 949–951
in a field theory of gravitation 326
six-vector **11**, 220, 237, 366–367, 423–427, 634, 636, 638, 645, 647, 663–664, 752, 878
Slebodzinski, Wladislaw 876
Smith, Henry J. S. 211
Smoluchowski, Marian von 836
solar system 184, 327, 362, 364, 377, 585–586, 1048
Sommerfeld, Arnold 10, 193, 212, 423–424, 787, 792, 836, 873, 909
four-dimensional formalism 235–243, 287
Sommerfeld-Laue notation 242
sound, propagation of 361
space 29, 31, 58, 287, 1079–1080
absolute 31, 35, 42, 131, 145, 161, 570, 572–573, 578, 582–583, 588, 590–592, 594, 598, 610, 617, 815, 1107
curvature of 162–164, 174
elliptic 163–164, 173
Euclidean 200, 218, 593, 1043, 1080, 1089, 1098
finiteness of 164
hyperbolic 162–163
non-Euclidean 1080
pseudo-Euclidean 200, 210
spherical 163
tangent 1063
topology of 165
spacetime 276, 279, 399, 874, 1041, 1056, 1089
axiomatic definition of 842
coincidence 53, 570
conformally flat 15, 414–415
diagram 193, 226, 228, 243–244
foliation 1050
kinematical structure of 1042
mechanics 193, 221, 225, 232, 239, 244
sickle **276**–280
singularity 325
thread **275**–277, 281–283
spacetime vector **1112**
equivalent **1112**
type I and II 424
special theory of relativity 366, 399, 569, 582, 586, 590, 594, 598–599, 605–610, 788, 831, 864, 934, 964, 1055, 1089
astronomy and 175–176
equations of motion in 280
Galilean frames in 1124
generalization of 552–565
gravitation and 9, 27, 362, 366, 390, 394, 396, 413, 490, 515, 1041
Lagrangian and gravitational mass 387, 389–390
non-Euclidean approach to 236–237
revision of 319
rigidity and 839
validity of 544–545
speed of light, *see* velocity of light
stability theory 786
star 33, 149, 162–163, 171, 184–185, 326, 344–345, 398
mass 345
maximal size 329
Stark, Johannes
Jahrbuch der Radioaktivität und Elektronik 417
statics 797
statistical mechanics 815
Stern, Otto 849
Stevin, Simon 801
stress 295, 523
elastic 499, 504, 509, 530, 535
fictitious gravitational 495, 518, 532
Maxwellian 531
relative 503, 507, 527
spatial 500, 525
tangential 492, 510
tensor, elastic 391, **501**, 504, 510–511, 524–525
total 447, 461, 534
see also elasticity theory
stressed body 437, 510
Strutt, John William, *see* Rayleigh, Lord
sufficient reason, principle of 594
Sun 294, 349
eclipse of 51, 167, 349
mass of 345
surface, developable 1067

T
Tait, Peter Guthrie 228, 573, 579–582, 585, 587, 589, 596, 787
telescope, water-filled 253
tensor calculus 195, 1057, 1093, 1097
thermodynamics 3, 5, 7, 25, 32, 786, 813, 817, 819–820
  phenomenological 591, 598
  second law 824, 834
  third law 827, 835
Thomson, James 573, 579
Thomson, Joseph John 788, 805
Thomson, William (Baron Kelvin) 4, 573, 787, 805
three-body problem 580, 771
tidal force 141, 1049, 1053
tides
  explanation of 141
time 31, 273, 287, 578, 581, 590, 799
  absolute 308, 572–573, 578, 586, 815, 945, 1025–1026, 1047, 1107
  as fourth dimension 1042
  direction of 798
  ephemeris 586
  measurement and inertio-gravitational field 1052
  proper 275, 289, 502, 541, 939–940, 1018–1019, 1125
  see also duration, simultaneity
time scale, inertial 588
Toepell, Michael 770
Tolman, Richard C. 59, 437, 840
topology, *see* analysis situs
torque 504, 1118, 1122–1123
torsion balance, *see* balance, torsion
torsion tensor 1075
translation, as defined by Weyl 1098
translations and rotations 1061
Trouton-Noble condenser 438
twin paradox 1057

U
umbrella model 3–4
unified field theory 57–59, 619, 858, 862, 959, 962–963, 966
  worldview 931
universe 585–586
  closed, as solution of Einstein's equations 173
  closed, considered by Schwarzschild 164
  large-scale structure 55
  spatially closed 601
  static 60

V
Van Dantzig, David 876
variational calculus 771, 777, 865, 874, 901, 936, 1001, 1015
  parameter invariance 936
variational principle 746–749, 869, 874, 887, 894, 912, 919–920, 922, 928, 949, 958, 962, 966
  for non-holonomic systems 806
Varičak, Vladimir 193, 236
Veblen, Oswald 770
vector 794–795, 1091
  sliding **1117**
  space-like 1050
  time-like 1050
  see also energy-momentum vector, four-vector, orthonormal vectors, six-vector, spacetime vector
vector addition 794, 796–797, 817
velocity 569, 1094
  distribution 824, 826
  four-vector 202, 204, 217–219, 224, 227, 230, 706
  infinitely large 288
velocity of light 253–255, 269, 271, 274, 306, 308, 310, 314, 323, 332–333, 341, 348, 350, 386, 396–398, 429, 490, 515, 536, 555, 606–607, 814, 821
  dependence on gravitational potential 12, 489, 695
Vermeil, Hermann 986–987
Veronese, Giuseppe 761
Villard, Paul 209
vis viva, *see* energy, kinetic
Voigt, Woldemar 201, 288
Volkmann, Paul 772, 816, 818
volume element 275, 1092
vortex theory 572
Voss, Aurel 145, 789, 793, 799, 816

W
Wacker, Fritz 208
Waerden, Bartel Leendert van der 864
Weber, Heinrich 211, 1087
Weber, Wilhelm 10, 40, 194
Weber's law 40, 95
Weierstrass, Karl 211–212, 772, 781
Weinstein, Max B. 225, 243
Weitzenböck, Roland 960–961
Weyl, Hermann 472, 631, 764, 819, 860, 862, 932, 949, 957, 959–963, 1044–1045, 1089–1105
  world metric 1090
Weylmann, Hermann 1042
Wiechert, Emil 194, 213, 776, 788, 791–792
Wien, Wilhelm 25, 241, 315, 423, 787–789, 809
Wilken, Alexander 213
Wisniewski, Felix Joachim de 230
world function, *see* Lagrangian, Mie's theory, Lagrangian
world line 217, 220, 223, 225–227, 232–233, 239–240, 244, 274–277, 279–280, 282–283, 308, 418, 451, 940–941, 1033, 1035
world matrix, *see* energy-momentum tensor
world parameter 922, 989–990, 995, 1003
world postulate 821
world tensor 366–367
world tube 1114
world, empty 1091
Wundt, Wilhelm 136

Z
Zangger, Heinrich 312–313, 423, 911
Zeeman effect 194, 792
Zeeman, Pieter 788
Zenneck, Jonathan 77, 112
  on gravitation 77–112
Zermelo, Ernst 214, 779
Zöllner, Karl Friedrich 8