5

Weyl's Classic, 1929

Although not fully appreciated at the time, Weyl's 1929 paper has turned out to be one of the seminal papers of the century, both from the philosophical and from the technical point of view.

From the philosophical point of view, the paper marked the completion of his 1918 ideas. He had always been convinced that there was a close analogy between gravitation and electromagnetism and was particularly impressed by the resemblance between the derivations of charge conservation and energy-momentum conservation in the respective theories. In this paper he was able to formulate the analogies between the two theories explicitly by means of the tetrad formalism and was able to overcome the objection to his 1918 theory by adopting London's reinterpretion of the non-integrable scale factor of the metric as a non-integrable phase factor of the wave function. But whereas London's formulation was tentative Weyl's formulation was complete and went beyond all previous ideas in proposing that electromagnetism be *derived* from the gauge principle. This proposal, which is something of a luxury in the case of electromagnetism, since that theory was already well-established, has turned out to be a powerful principle for deriving the nuclear interactions and to be the common principle underlying all the known fundamental interactions.

From the technical point of view, Weyl introduced in this paper many of the concepts in field theory that are now taken for granted. This can be seen by considering the six sections of the paper in turn. In the first section, he introduced the two-component spinor theory in Minkowski space and discussed the validity of parity and time-reversal invariance P and T. In this and the following section, he developed the first systematic formulation of what is now called the tetrad, or Vierbein, formalism. The aim was not only to formulate spinor theory for curved spaces but to give a systematic derivation of the Noether conservation laws and to make the resemblance between gravitation and electromagnetism manifest. In the third section, he continued the formulation of spinor theory for curved manifolds by introducing the spin connection and constructing an invariant action. In the fourth section, he derived the energy-momentum conservation laws from invariance with respect to both general coordinate transformations and Lorentz transformations of the tetrad (Noether's theorem). He then recast gravitational theory in the tetrad formalism (section five) with a view to exhibiting the analogies between it and electromagnetism. In the final section, he came to what he considered the most fundamental part of the paper, namely, the derivation of electromagnetic theory from the gauge principle.

The debut of the paper was inauspicious. Before publishing the full paper, Weyl published a short summary in the *Proceedings of the U.S. National Academy* (while on a visit to Princeton), and this resume was immediately attacked by Pauli [1], who wrote to him as follows:

> Before me lies the April edition of the *Proc. Nat. Acad (U.S.).* Not only does it contain an article from you under "Physics" but shows that you are now in a "Physical Laboratory": from what I hear you have even been given a chair in "Physics" in America. I admire your courage; since the conclusion is inevitable that you wish to be judged, not for successes in pure mathematics, but for your true but unhappy love for physics.

Pauli's objection was partly to the intrusion of a pure mathematician into physics, especially in view of the failure of the 1918 theory, and partly due to his well-known dislike [1] of parity violation and hence of 2-component spinors. The strength of his dislike can be judged from his dismissal of the 2-component theory in the 1933 *Handbuch* article cited in Chapter 1 and from his famous reaction to Salam's proposal to use the 2-component theory for the neutrino after parity violation had been proposed in 1955: "Give my greetings to my friend Salam and tell him to think of something better."

On reading the full version of Weyl's paper, however, Pauli's attitude softened [1], and he wrote a second letter to say:

> In contrast to the nasty things I said, the essential part of my last letter has since been overtaken, particularly by your paper in *Z. f. Physik.* For this reason I have afterward even regretted that I wrote to you. After studying your paper I believe that I have really understood what you wanted to do (this was not the case in respect of the little note in the Proc. Nat. Acad.) First let me emphasize that side of the matter concerning which I am in full agreement with you: your incorporation of spinor theory into gravitational theory. I am as dissatisfied as you are with distant parallelism and your proposal to let the tetrads rotate independently at different space-points is a true solution. Your method is valid even for the massive [Dirac] case. I thereby come to the other side of the matter, namely the unsolved difficulties of the Dirac theory (two signs of m_0) and the question of the 2-component theory. In my opinion these problems will not be solved by gravitation ... the gravitational effects will always be much too small.

In fact, Pauli accepted "what Weyl wanted to do" to such an extent that he incorporated most of it into the famous *Handbuch* article on quantum theory quoted earlier. Indeed, he incorporated the Noether theorem and the gauge principle to such an extent that many physicists who had studied the *Handbuch* article but had not read Weyl's paper attributed the introduction of these concepts to Pauli. For example C. N. Yang [2] writes:

> I studied Pauli's Handbuch article thoroughly... and I was very much impressed with the idea that charge conservation was related to the invariance of the theory under phase changes and even more impressed with the fact that gauge-invariance determined all the electromagnetic interactions, ideas which I later found out were originally due to Weyl.

As we shall see later, by 1953 Pauli had become an ardent proponent of Weyl's point of view regarding the gauge principle.

Weyl's 1929 paper is reproduced here and the contents are displayed in the accompanying abstract. As mentioned above it consists of six separate sections and each one is worth some comment.

WEYL'S §1. TWO-COMPONENT THEORY

In the first part of section 1, Weyl introduced the concept of two-component spinors and discussed the question of P and T invariance for 2-spinors in Minkowski space.

He began by defining a 2-component spinor as a vector in the 2-dimensional (self-) representation of the group $SL(2, C)$,

$$\phi \quad \rightarrow \quad \phi' = S(l)\phi \quad \text{where} \quad S(l) = e^{\omega_{ab}\sigma_{ab}} \in SL(2, C), \tag{105}$$

and the ω's and σ's are the parameters and generators of $SL(2, C)$, respectively. He then pointed out that the bilinears

$$v_a = \phi^\dagger \sigma_a \phi \qquad \sigma_a = (1, \vec{\sigma}), \tag{106}$$

where $\vec{\sigma}$ are the Pauli matrices and the dagger denotes hermitian conjugation, transform according to the 4-vector representation of $SL(2, C)$ and that this representation of $SL(2, C)$ is isomorphic to the connected part of the Lorentz group. Thus the vectors v_a are Lorentz 4-vectors. Weyl noted that there was no bilinear Lorentz *scalar* and concluded that 2-component spinors must be massless.

The mapping of an $SL(2, C)$ spinor into a Lorentz vector exhibited in (106) shows that $SL(2, C)$ is the (double) covering group for the connected Lorentz group, and with this in mind the generators σ of $SL(2, C)$ are often written as

$$\sigma_{ij} = i\epsilon_{ijk}\sigma_k \quad \text{and} \quad \sigma_{io} = \sigma_i \quad i, j = 1, 2, 3, \tag{107}$$

the σ_i being the Pauli matrices, where the σ_{ij} and σ_{io} correspond to rotations and "pure" Lorentz transformations respectively, The complex conjugate field ϕ^* belongs to the complex conjugate representation of $SL(2, C)$. This representation is inequivalent for the full $SL(2, C)$ group but equivalent for the rotation subgroup and in order to make the rotations the same for the field and its conjugate it is usual to use not the simple complex conjugate ϕ^* but $\bar{\phi} = i\sigma_2\phi^*$.

Weyl pointed out that since the fourth component v_o of the vector v_a acquires a factor $S^\dagger S > 0$ under an $SL(2, C)$ transformation, the $SL(2, C)$ transformation excludes time-reversal T. He considered this to be an advantage since it takes account of the fact that nature is not time-reversable. He next pointed out that since a parity transformation P changes the signs of the pure Lorentz generators σ_{io}, and thereby converts the self-representation of $SL(2, C)$ to its *in*equivalent contragredient representation, the 2-component theory also excludes the linear implementation of parity. Since this was long before the discovery of parity violation, he considered this to be a disadvantage of the theory. Indeed, since he thought that the mass problem would eventually solved by gravitation, he attributed the need for a Dirac 4-component theory to parity rather than to mass.

It is remarkable that Weyl should even consider the possibility of time reversal and parity violation at this time. In fact, Weyl not only considered these possibilities but in the edition of his RZM book published shortly afterwards made the statement: "The problem of the proton and electron will be mixed with the symmetry properties of the quantum theory with respect to interchange of left and right, past and future, and positive and negative charge." Thus (modulo the confusion between the proton and the anti-electron that was prevalant before the experimental discovery of the positron) he not only foreshadowed the later developments in P and T violation but foreshadowed the CPT theorem. All this was at a time when, as Yang [2] put it, "[N]obody, absolutely nobody, was in any way suspecting that these symmetries were related in any manner. It was only in the 1950's that the deep connection between them was discovered.... What had prompted Weyl in 1930 to write the above passage is a mystery to me."

As Weyl spinors are less familiar than Dirac spinors, it may be useful to place them in the Dirac context. The conventional basis for the Dirac spinors is the one in which the Dirac matrix γ_o is diagonal, because γ_o serves both as parity operator and mass operator in the 4-dimensional Dirac space. However, an alternative basis is one in which the Dirac matrix γ_5 is diagonal, and this basis is very natural from the Lorentz point of view because γ_5 is Lorentz-invariant. In the γ_5 basis, the Dirac representation of the Lorentz group decomposes into two inequivalent irreducible components, labeled by $\gamma_5 = \pm 1$, respectively. In this basis, the Dirac matrices γ_μ are, of course, strictly off-diagonal because they anti-commute with γ_5. The Weyl spinors are then identified as the vectors in the 2-component irreducible representations. More precisely, they are identi fied as

$$\phi = \frac{1}{2}(1+\gamma_5)\psi \quad \text{and} \quad \xi = \frac{\sigma_2}{2i}(1-\gamma_5)\psi^*. \tag{108}$$

From this point of view, the masslessness and parity violation of the Weyl spinors are simply due to the fact that γ_5 and γ_o do not commute.

Free Weyl spinors satisfy the equation

$$\vec{\sigma}.\vec{p}\psi = \pm |p|\,\psi, \tag{109}$$

where the $\pm$ refers to the eigenvalue of γ_5. Thus for such spinors γ_5 coincides with the helicity, which is defined [3] as $h = \vec{\sigma}.\vec{p}/p$. As the eigenvalue γ_5 therefore determines whether the spin is parallel or anti-parallel to the 3-momentum, it is called the *chirality*, or handedness, and Weyl spinors are said to be *chiral.*

It was only when parity was found to be violated in 1956 that the importance of chirality came to be appreciated. The subsequent discovery that parity was not only violated, but *maximally* violated [4], in the sense that the fundamental currents had the $V-A$ structure $j_\mu = \bar{\psi}\gamma_\mu(1-\gamma_5)\psi$, was simply the discovery that these currents are chiral.

The $V-A$ discovery [5] has led to a picture of the fundamental interactions in which chiral symmetry is regarded as fundamental. The idea [6] is that the fundamental interactions are chiral in their symmetric phase and the observed non-chiral phenomenology is due to a spontaneous breakdown of this symmetry. Remnants of the original chiral symmetry are still to be observed in the $V-A$ structure of the weak currents and the chirality of the neutrinos (to within experimental error). There are good astrophysical reasons [6] for assuming that the early universe was chirally symmetric and that the spontaneous breakdown of chirality was caused by the cooling of the universe due to expansion. If this is the correct picture, then Nature in its symmetric phase prefers Weyl spinors to Dirac spinors!

It is perhaps worth noting from (108) that a Dirac spinor is equivalent to a single Weyl spinor ($\phi = \xi$) if and only if the Dirac spinor is real in the sense that $\psi^* = i\sigma_2\gamma_0\psi$. In that case the Dirac spinor is called a Majorana spinor [7] and admits a real basis. Thus a Weyl spinor is completely equivalent to a Majorana spinor, each having four real components. Interestingly enough, parity can be implemented linearly on Majorana spinors because

$$\phi(\vec{p}) \to \phi^*(-\vec{p}) \qquad \Leftrightarrow \qquad \psi(\vec{p}) \to \gamma_o\psi(-\vec{p}). \tag{110}$$

However, the Weyl version of (110) shows that it is really a CP transformation [5]. The CP transformation coincides with the P transformation for Majorana spinors because they are real and hence have $C = 1$.

WEYL'S §2. TETRAD FORMALISM

Weyl wished to incorporate his 2-component spinor theory into gravitational theory. It was already known that the use of spinors in gravitational theory presents new problems because, in contrast to the Lorentz group, the group of general coordinate transformations (diffeomorphisms) does not admit a 2-1 covering group. However, as had recently been pointed out by Wigner [8], the solution to this problem is to use local tetrads, a concept which had been introduced by Einstein [9] for a different purpose shortly before.

The simplest way to understand Wigner's argument is to note that from the uniqueness (up to conjugation) of the irreducible representation of the Dirac matrices, the only irreducible representation of the curved-space Dirac matrix equation

$$\{\gamma_\mu(x), \gamma_\nu(x)\} = 2g_{\mu\nu}(x) \tag{111}$$

is

$$\gamma_\mu(x) = e^a_\mu(x)\gamma_a, \tag{112}$$

where the γ_a are the flat-space Dirac matrices and the coefficients $e^a_\nu(x)$ satisfy the condition

$$\eta_{ab}e^a_\mu(x)e^b_\nu(x) = g_{\mu\nu}(x), \tag{113}$$

where η is a rigid Minkowski metric. But (113) is just the condition that the $e^a_\mu(x)$ form local tetrads [10]. Thus the solution of the Dirac matrix equation on curved spaces requires local tetrads.

Since the same kind of argument applies to any kind of spinors, Weyl was forced to introduce the concept of tetrads in order to incorporate his 2-component spinors into gravitational theory. Accordingly, in section 2, he proceeded to develop the tetrad formalism in a systematic manner. The formalism then turned out to be useful not only for handling spinors on curved spaces, but for other purposes, such as deriving the energy-momentum conservation laws and making the analogy between gravitation and electromagnetism manifest. Although Weyl introduced factors of i in order to work in Euclidean space, this is not necesssary, and for our brief summary, we shall use the Minkowski space formulation.

It is clear from (113) that every tetrad determines a Riemannian metric uniquely. But it is also clear that the converse cannot be true since the tetrad has 16 independent components and the metric has only 10. The 6 degrees of freedom left undetermined by the metric may be characterized precisely by the elements of a 6-parameter internal Lorentz group $\mathcal{L}$. To see this, let

$$\hat{e}^a_\mu(x) = h^a_b(x)e^b_\mu(x) \tag{114}$$

be any other tetrad satisfying (113). Then, from the definition of the tetrad we have

$$h(x)^a_c \eta_{ab} h^b_d(x) = \eta_{cd} \quad \text{or} \quad h^t \eta h = \eta, \tag{115}$$

where superscript t denotes transpose. But this is just the condition that $h(x) \in \mathcal{L}$. So the tetrads are determined by the metrics up to local $\mathcal{L}$ transformations.

If the elements $h(x) \in \mathcal{L}_c$, where $\mathcal{L}_c$ denotes the connected part of the local Lorentz group, are be written in the form

$$h(x) = e^{\omega_{ab}(x)\Sigma_{ab}}, \tag{115a}$$

where the $\omega_{ab}(x)$ are the parameters and the Σ's are the generators of $\mathcal{L}_c$, then the $\mathcal{L}_c$ transformations,

$$e^a_\mu(x) \quad \rightarrow \quad R^a_b(x) e^b_\mu(x) \quad \text{where} \quad R(x) = e^{\omega_{cd}(x)\Sigma_{cd}}, \tag{116}$$

may be thought of as a non-abelian generalization of the EM gauge transformations in Chapter 1 since the parameters are x-dependent and the transformations leave physical quantities (in this case, the components of the Riemannian metric) invariant. Thus at this point Weyl had already revealed one of the analogies between electromagnetism and gravitation.

Once the tetrad has been defined by (113) it can be used to convert any Riemannian tensor $T_{ij\ldots k}(x)$ to an internal tensor $\Theta_{ab\ldots c}$ according to the rule

$$\Theta_{ab\ldots c}(x) = e^i_a(x) e^j_b(x) \ldots e^k_c(x) T_{ij\ldots k}(x). \tag{117}$$

The quantities T transform as Riemannian tensors and internal-Lorentz scalars, and the quantities Θ transform the other way around.

Weyl then went on to define an internal version of the connection according to

$$A^b_{\mu c}(x) = e^b_\sigma \left(\nabla_\mu(\Gamma) e\right)^\sigma_c \equiv \Gamma^b_{\mu c} + J^b_{\mu c}, \tag{118}$$

where J_μ is the current

$$J^b_{\mu c} = e^b_\xi \partial_\mu e^\xi_c = \left(e^{-1}\partial_\mu e\right)^b_c. \tag{119}$$

In the last expression in (119), e^{-1} is to be understood as the matrix inverse, where e is considered to be an element of $GL(4, R)$, and from this one sees that the current J_μ lies in the $SL(4, R)$ Lie algebra. Note that in the flat-space limit $\Gamma = 0$ and $e \in \mathcal{L}$, the internal connection A_μ does not reduce to zero but to the current $e^{-1}\partial_\mu e$ for $e \in \mathcal{L}$. This shows that A_μ is not completely metrical. Equation (118) can, of course, be inverted to obtain Γ_μ in terms of A_μ.

Since it is defined by the Riemannian covariant derivative, it is clear that A_μ transforms covariantly with respect to the diffeomorphism group $\mathcal{D}$. On the other hand, it transforms non-covariantly with respect to $\mathcal{L}$,

$$A_\mu(x) \quad \to \quad A^h_\mu(x) = h^{-1}(x)A_\mu(x)h(x) + h^{-1}(x)\partial_\mu h(x) \qquad h(x) \in \mathcal{L}, \tag{120}$$

the inhomogeneous term coming from the current-term J_μ in A_μ. Thus by replacing Γ_μ by A_μ one has replaced a connection with respect to $\mathcal{D}$ by a connection with respect to $\mathcal{L}$. So what advantage has been gained?

The biggest advantage, perhaps, is that it exhibits very explicitly the fact that gravitation is a gauge theory. Indeed, the analogy between (120) and the electromagnetic gauge transformations is evident. Although Weyl did not know it at the time, the analogy between (120) and the gauge transformations of the connections in non-abelian gauge theory is even more striking. (Compare, for example, equation (120) with equation (3) of the Yang-Mills paper in Chapter 8). Indeed, apart from the fact that the Lorentz group is non-compact, whereas the non-abelian gauge groups are compact, the two sets of transformations are identical.

A more technical advantage is that much of the differential geometry can be converted into algebra. For example, the statement that the space-time metric $g_{\mu\nu}$ is covariantly constant with respect to Γ_μ converts to the statement that the internal metric $\eta_{\mu\nu}$ is covariantly constant with respect to A_μ, and since $\eta_{\mu\nu}$ has no derivative, this is just the statement that A_μ is an element of the Lie algebra L of $\mathcal{L}$,

$$\nabla_\lambda(\Gamma)g_{\mu\nu}(x) = o \quad \Leftrightarrow \quad A^c_{\mu a}\eta_{cb} + A^c_{\mu b}\eta_{ac} = 0 \quad \Leftrightarrow \quad A_\mu \in \mathrm{L}. \tag{121}$$

Note that this implies that $A_{\mu bc}$ is anti-symmetric in the indices b, c.

Since the metric is a function of the tetrad it is possible to eliminate the metric from the definition of A_μ and thus express A_μ in terms of the tetrad alone. From (118) a short computation shows that A_μ is actually a linear function of the currents (119) with tetrad coefficients, namely, $A_{\mu[bc]} = e^a_\mu A_{a[bc]}$ where

$$2A_{a[bc]} = J_{abc} - J_{acb} + J_{bac} + J_{bca} - J_{cab} - J_{cba}. \tag{122}$$

This is the equation following (11) in Weyl's paper. It verifies that $A_{a[bc]}$ and hence $A_{\mu[bc]}$ is anti-symmetric in b and c. The generalization of (122) in the presence of torsion is simply

$$\tilde{A}_{\mu[bc]} = A_{\mu[bc]} + \theta_{\mu[bc]} \tag{123}$$

where θ is the contorsion tensor and was used by Utiyama (p. 231).

WEYL'S §3. SPINORS IN CURVED SPACE

From the group-theoretical point of view, the introduction of the local tetrads extends the group $\mathcal{D}$ of general coordinate transformations (diffeomorphisms), which lies at the heart of gravitational theory, to the semi-direct product group $\mathcal{D} \wedge \mathcal{L}$, where $\mathcal{L}$ is the group of local Lorentz transformations, and Weyl's first application of the tetrad formalism was to exploit this extension to solve the problem of incorporating 2-spinors into gravitational theory.

Accordingly (at the end of section 1) he defined 2-spinors $\phi_\alpha(x)$ in curved space to be 2-spinors with respect to $\mathcal{L}$, whose components were scalars with respect to diffeomorphisms. He concluded the section by giving their infinitesimal transformations with respect to $\mathcal{L}$, namely,

$$\delta_{\mathcal{L}}\phi^\alpha(x) = \frac{1}{2}\omega_{ab}(x)(\sigma_{ab})^\alpha_\beta\phi^\beta \qquad \alpha, \beta = 1, 2, \tag{124}$$

where the $\omega_{ab}(x)$ are the local parameters of $\mathcal{L}$ and the σ's are the the rigid generators of $SL(2, C)$ defined following (106). (This is equivalent to the second equation after (6) in Weyl's paper, with ω and σ identified as do and A, respectively, and with a, b instead of α, β). As might be expected, the definition (124) implies that

$$\psi^\dagger(x)\sigma_\mu(x)\psi(x), \tag{125}$$

which is invariant with respect to $\mathcal{L}$, is a Riemannian vector. The expression (125) is evidently the analogue for Weyl spinors of the Dirac vector $\bar{\psi}\gamma_\mu\psi$.

In section 3, Weyl went on to define the covariant derivative for spinors and construct the action for free spinors in a gravitational field. By considering parallel transport and using the fact that the covariant derivative for vectors can be written as

$$D_\mu = \partial_\mu + A^a_{\mu b}\Sigma^b_a, \quad \text{where} \quad (\Sigma_{ab})^{cd} = \frac{1}{2}\left(\delta^c_a\delta^d_b - \delta^c_b\delta^d_a\right) \tag{126}$$

are the generators of $\mathcal{L}$, he showed that the covariant derivative for the spinors must be of the similar form

$$D_\mu = \partial_\mu + A^a_{\mu b}\sigma^b_a, \tag{127}$$

where the σ^b_a are the generators of the spinor representation of $SL(2, C)$. Note that the σ's have vector labels to match the A_μ's but act on spinors. As might be expected, there is a corresponding formula for general spin, namely, $D_\mu = \partial_\mu + A^a_{\mu b}\tau^b_c$, where the τ's are the generators of the relevant representation of the internal $SL(2, C)$. The quantity $S_\mu = A^a_{\mu b}\tau^b_c$ is now called the spin connection.

Using the covariant derivative, Weyl then constructed the invariant action for the 2-spinor in a curved space, namely,

$$\begin{aligned} \mathrm{m} &= \frac{1}{i}\int \psi^* \left(\sigma^\mu(x) D_\mu + D_\mu \sigma^\mu(x)\right)\psi \\ &= \frac{1}{i}\int \psi^* \left(\sigma^\mu(x) D_\mu + \frac{1}{2}\sigma_a \frac{\partial e^a_\mu(x)}{\partial x_\mu}\right)\psi, \end{aligned} \tag{128}$$

where $\sigma_\mu(x) \equiv e^a_\mu(x)\sigma_a$ and the σ_a are the unit and Pauli matrices of (106). The symmetrization in (128) guarantees hermiticity (up to a total divergence) and current conservation.

WEYL'S §4. THE NOETHER THEOREM

The next application of the tetrad formalism was to derive the conservation of linear and angular momentum using invariance with respect to (i) coordinate transformations and (ii) internal Lorentz transformations of the tetrad. The conservation laws for energy and momentum in flat space had, of course, been derived much earlier by Hilbert and had been derived for gravitational theory by Einstein and by Weyl himself in his 1918 paper. Furthermore, in the paper discussed in Chapter 1, Noether had given the formal prescription for obtaining the conserved current associated with *any* continuous symmetry of a Lagrangian. Thus, although Weyl did not refer to Noether, this particular result of his was actually a special case of Noether's theorem. Weyl's contribution was to make the result familiar to physicists in the context of field theory, to derive it using the tetrad formalism, and to exhibit the analogy between the energy-momentum and electromagnetic conservation laws.

As discussed in Chapter 1, the energy-momentum tensor density $T_{\mu\nu}(x)$ derived in the straightforward Noether manner is not always symmetric, although the symmetry of $T_{\mu\nu}(x)$ is necesssary for the conservation of angular momentum. One advantage of the tetrad formalism is that, as shown quite simply by Weyl, the energy-momentum tensor density $T_{\mu\nu}(x)$ derived by imposing the tetrad invariance (i) is automatically symmetric and thus the corresponding angular momentum is conserved.

Although the analogy between the energy-momentum and electric-charge conservation laws was one of the features of his theory that convinced Weyl that he was on the right track, it should be recalled that the conservation laws are derived from the *rigid*, i.e., space-time–independent, part of the group. The space-time–dependent part is used only to motivate the introduction of the gauge fields. Thus the gauge fields and the conservation laws are actually associated with complementary parts of the gauge groups, namely, the local subgroups $\tilde{G}(x)$, where $\tilde{G}(0) = 1$ and the rigid quotient groups $G(0)$, respectively.

Weyl's use of the Noether theorem was restricted to classical fields, but, as mentioned in Chapter 1, the classical conservation laws generalize to quantum field theory, either as Ward identities or as symmetries broken by anomalies, and it is fitting that one of the most interesting anomalies at the present time is the conformal anomaly, which is due to the breakdown at the quantum-mechanical level of the Weyl symmetry $g_{\mu\nu}(x) \to e^{\sigma(x)} g_{\mu\nu}(x)$, where $\sigma(x)$ is a scalar parameter.

WEYL'S §5. GRAVITATION

In the section on gravitation Weyl introduced no new physics but reformulated the usual theory in the tetrad formalism. In particular, he derived the conservation laws and expressed the Riemann tensor in the tetrad form

$$R^{a}_{\mu\nu b} = [D_\mu, D_\nu]^a_b = \partial_\mu A^a_{\nu b} - \partial_\nu A^a_{\mu b} + A^a_{\mu c} A^c_{\nu b} - A^a_{\nu c} A^c_{\mu b}, \tag{129}$$

or in modern compact notation, $R = \partial \wedge A + [A, A]$. He thus exhibited the analogy between the Riemann tensor and the electromagnetic field strengths and foreshadowed the definition of field strengths in non-abelian gauge theory. He also derived the Einstein action in the tetrad form

$$\int \sqrt{-g} R = \int \sqrt{-g} \left(A_{a[bc]} A^{c[ab]} + A_c A^c \right) \quad \text{where} \quad A_c \equiv A^a_{[ac]}, \tag{130}$$

where, as in the space-time formulation, a total divergence has been dropped on the right-hand side in order to obtain an expression which has no second derivatives. The formula (130) is particularly useful for dimensional reduction.

As mentioned earlier, it is clear from remarks made by Weyl in this and other works that he regarded the analogy between gravitation and electricity as profound, and both the conservation laws and the tetrad formulation emphasized this analogy.

WEYL'S §6. ELECTROMAGNETISM

In the final section, Weyl came to what he considered to be the crux of the paper, namely, the derivation of electromagnetism from the gauge principle.

The first step was to note that his two-component theory (and by extension, any spinorial theory) had an intrinsic abelian gauge freedom. The point was that the $SL(2, C)$ group in the $SL(2, C) \to \mathcal{L}_c$ homomorphism by which spinors are defined does not include the full group of linear transformations on spinor space. The latter group is $GL(2, C)$ and differs from $SL(2, C)$ by the central subgroup zI, where z is any complex number and I is the unit 2×2 matrix. Unitarity reduces z to a phase factor, $z = e^{i\alpha}$ for real α, but there is no way to reduce it further. Thus there is an intrinsic phase freedom in the spinor theory, namely,

$$\psi(x) \to e^{i\alpha}\psi(x), \tag{131}$$

where α is the group parameter.

The second step in Weyl's argument was to note that, just as it is natural to generalize the rigid Minkowski tetrad to a local tetrad, it is natural to generalize the phase exponent α to a local function, in which case the spinorial phase transformation (131) generalizes to

$$\psi(x) \to e^{i\alpha(x)}\psi(x). \tag{132}$$

Furthermore, since the exponent $\alpha(x)$ is independent of the tetrad, its locality should be an intrinsic property which persists even in the case of rigid tetrads. In other words, the curvature is used only to *motivate* the locality of the phase transformation and the latter may remain local even when the curvature vanishes.

The third step in Weyl's argument was to assign the known fermionic fields to representations of the phase group according to

$$\psi_e(x) \to e^{ie\alpha(x)}\psi_e(x), \tag{133}$$

where e is the electric charge, and to note that just as diffeomorphic invariance requires that $\partial_\mu \to \Delta_\mu = \partial_\mu + \Gamma_\mu(x)$, so invariance with respect to the abelian gauge group (133) requires that $\Delta_\mu \to D_\mu = \Delta_\mu + \frac{ie}{\hbar c}A_\mu(x)$, where the field $A_\mu(x)$ is a connection for the abelian group. Of course, since there is only one parameter $\alpha(x)$, the field $A_\mu(x)$ has only one component. In the flat-space limit the covariant derivative simplifies to $D_\mu = \partial_\mu + \frac{ie}{\hbar c}A_\mu$.

The electrodynamics was then obtained by applying the gauge principle, i.e., taking any Lagrangian density $L(\psi(x), \partial\psi(x))$, which is invariant with respect to the rigid subgroup of the phase group in order to conserve electric charge, and making it invariant with respect to the local phase group by changing all the derivatives to covariant ones,

$$L(\psi(x), \partial\psi(x)) \quad \to \quad L(\psi(x), D\psi(x)). \tag{134}$$

Of course, a kinetic term for the gauge field $A_\mu(x)$ itself has to be added, since otherwise the field $A(x)$ is reduced to a Lagrange multiplier and the theory is constrained. But gauge invariance, Lorentz invariance and the usual restrictions on the order and power of the derivatives restrict the kinetic term to multiples of $F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x)$, and this term, which had already appeared in the 1918 paper, was duly added by Weyl. With this prescription, the formulation of the gauge principle was complete.

Today the second and third steps of Weyl's proposal are fully accepted. Indeed, their scope has been widened to include non-abelian gauge groups. The first step, however, is no longer valid and serves only a historical motivation for the other two. Weyl had identified the phase group in (132) as the electromagnetic gauge group because in 1929 the only known fermions (the electron and the proton) were electrically charged and the only known boson (the photon) was not. However, once uncharged fermions (such as the neutron and neutrino) and charged bosons (such as the $\pi^{\pm}$ mesons) had been discovered experimentally, this identification became untenable, and today we identify the charge associated with (131) as fermion number rather than the electric charge. Thus, as in the case of the Dirac equation, the motivation was a little weak, but the final result, guided by the intrinsic beauty of the mathematics, was profound.

Up to now we have spoken of the analogies between electromagnetism and gravitation but the gauge principle serves to emphasize also the *differences*. The most fundamental difference, of course, is that there is no equivalence principle, or universal coupling, for electromagnetism. This manifests itself in the fact that equation (132) is not sufficient but must be replaced by equation (133) in which the electric charge is introduced explicitly. Thus the way in which matter interacts with the electromagnetic field is not uniform but is determined by the charge. From the mathematical point of view, the charge is actually the *character* of the representation of the gauge group to which the matter belongs, and we shall see later that the property of the interaction being determined by the group representation is common to all gauge theories. This is particularly important in non-abelian theories, for which the representations are no longer one-dimensional. The other differences between gravitation and electromagnetism stem from the fact that whereas gravitation uses the metric, the connection and the field strength (g, Γ, R), electromagnetism uses only the connection and the field strengths (A, F). Finally, the gravitational Lagrangian density R is linear in the field strengths, but the electromagnetic Lagrangian density F^2 is quadratic.

As we have already seen in Chapter 4, the substitution $\partial \to \partial + \frac{ie}{\hbar}A$ is equivalent to the substitution

$$\psi(x) \quad \to \quad e^{\frac{ie}{\hbar}\int^x A_\mu(y)dy^\mu}\psi(x), \tag{135}$$

and thus Weyl was led once more to the non-integrable factor of his 1918 paper, but applied now to the wave function and with a factor i, as had been suggested by Schrödinger and London. Thus the wheel had turned full cycle both for Weyl's physical theory and for his philosophy. As he said himself:

> This new principle of gauge-invariance, that derives not from speculation but from experiment, shows that electromagnetism is an accompanying

> phenomenon, not of the gravitational field but of the wave-function represented by ψ.

The above theory of electromagnetism was very satisfactory to Weyl from a number of points of view. First, the theory of electromagnetism was derived from a geometrical principle but the geometrical derivation was in complete accord with the empirical one. This is the meaning of the statement just quoted to the effect that the the theory "derives not from speculation but from experiment." Second, the central feature of the theory was the covariant derivative, and in this respect the theory was similar to gravitation: As he says:

> Since gauge-invariance involves an arbitrary function $\alpha(x)$ it has the character of general relativity and can only be understood in this context.

The other similarities and dissimilarities with gravitation have already been discussed.

Even Pauli was impressed by these arguments [1] and wrote:

> Here I must admit your ability in Physics. Your earlier theory with $g'_{ik} = \lambda g_{ik}$ was pure mathematics and unphysical. Einstein was justified in criticising and scolding. Now the hour of your revenge has arrived.

As already mentioned, Pauli proceeded to incorporate many of Weyl's ideas into his *Handbuch* article and by 1953 he had become an ardent proponent of the gauge principle.

Electron and Gravitation

by Hermann Weyl in Princeton, N.J.

Zeit. f. Physik 330 56 (1929)

Introduction. Relationship of General Relativity to the quantum-theoretical field equations of the spinning electron: mass, gauge-invariance, distant-parallelism. Expected modifications of the Dirac theory. —I. Two-component theory: the wave-function ψ has only two components. —§ 1. Connection between the transformation of the ψ and the transformation of a normal tetrad in four-dimensional space. Asymmetry of past and future, of left and right. —§ 2. In General Relativity the metric at a given point is determined by a normal tetrad. Components of vectors relative to the tetrad and coordinates. Covariant differentiation of ψ. —§ 3. Generally invariant form of the Dirac action, characteristic for the wave-field of matter. —§ 4. The differential conservation law of energy and momentum and the symmetry of the energy-momentum tensor as a consequence of the double-invariance (1) with respect to coordinate transformations (2) with respect to rotation of the tetrad. Momentum and angular momentum for matter. —§ 5. Einstein's classical theory of gravitation in the new analytic formulation. Gravitational Energy. —§ 6. The electromagnetic field. From the arbitrariness of the gauge-factor in ψ appears the necessity of introducing the electromagnetic potential. Gauge-invariance and charge conservation. The space-integral of charge. The introduction of mass. Discussion and rejection of another possibility in which electromagnetism appears, not as an accompanying phenomenon of matter, but of gravitation.

Introduction

In this paper I develop in detailed form a theory which embraces gravitation, electromagnetism and matter and which was sketched in the Proc. Nat. Acad., April 1929. Many authors have noticed the connection between Einstein's theory of distant-parallelism and the spin theory of the electron [1]. In spite of certain formal agreements my Ansatz is distinguished by the fact that I dispense with distant-parallelism and stick to Einstein's classical theory of gravitation. The adaption of the Pauli-Dirac theory of the spinning electron to gravitational theory promises to lead to physically fruitful results for two reasons: 1. The Dirac theory, in which the wave-field of the electron is described by four components gives twice too many energy levels; one should therefore return to the two-component Pauli theory, without sacrificing relativistic invariance. But that is prevented by the fact that the electron mass appears in the Dirac action. Mass, however, is a gravitational effect; thus there is a hope of finding a substitute in the theory of gravitation that would produce the required corrections. 2. The Dirac field-equations for ψ together with the Maxwell equations for the four potentials f_p of the electromagnetic field have an invariance property which is formally similar to the one which I called gauge-invariance in my 1918 theory of gravitation and electromagnetism; the equations remain invariant when one makes the simultaneous substitutions

$$\psi \quad \text{by} \quad e^{i\lambda}\psi \qquad \text{and} \qquad f_p \quad \text{by} \quad f_p - \frac{\partial \lambda}{\partial x_p}$$

where λ is understood to be an arbitrary function of position in four-space. Here the factor $\frac{e}{ch}$, where $-e$ is the charge of the electron, c is the speed of light and $\frac{h}{2\pi}$ is the quantum of action, has been absorbed in f_p. The connection of this "gauge-invariance" to the conservation of electric charge has also not been explored. But a fundamental difference, which is important to obtain agreement with observation, is that the exponent of the factor multiplying ψ is not real but pure imaginary. ψ now plays the role that Einstein's ds played before. It seems to me that this new principle of gauge-invariance, which follows not from speculation but from experiment, tells us that the electromagnetic field is a necessary accompanying phenomenon, not of gravitation, but of the material wave-field represented by ψ. Since gauge-invariance involves an arbitrary function λ it has the character of "general" relativity and can naturally only be understood in that context.

I prefer not to believe in distant-parallelism for a number of reasons. First my mathematical intuition objects to accepting such an artificial geometry; I find it difficult to understand the force that would keep the the local tetrads at different points and in rotated positions in a rigid relationship. There are, I believe, two important physical reasons as well. The loosening of the rigid relationship between the tetrads at different points converts the gauge-factor $e^{i\lambda}$, which remains arbitrary with respect to ψ, from a constant to an arbitrary function of space-time. In other words, only through the loosening of the rigidity does the established gauge-invariance become understandable. Secondly the freedom of rotating the tetrads independently at different points is, as we shall see, equivalent to the symmetry and conservation of the energy-momentum tensor.

In every attempt to establish the quantum-theoretical field equations one must keep in mind that these are not to be compared directly with experiment but yield only, after quantization, the basis for statistical predictions concerning the behaviour of matter and light-quanta. The Maxwell-Dirac equations contain in their present form only the electromagnetic potentials f_p and the wave-function ψ of the electron. Without doubt the wave-function ψ' of the proton must also be included. Furthermore, in the field equations ψ, ψ' and f_p should be functions of the same space-time point, before quantization one should not require for example that ψ be a function of the space-time point (t, x, y, z) and ψ' be a function of an independent space-time point (t', x', y', z'). It is reasonable to expect that in the two component-pairs of the Dirac field one pair should correspond to the electron and the other to the proton. Furthermore there should appear two electrical conservation laws, which (after quantization) should state the separate conservation of the number of electrons and protons. These would have to correspond to a two-fold gauge-invariance, involving two arbitrary functions.

We first investigate the situation in special relativity to see if, and to what extent, the formal requirements of group theory, independently of the field-equations that are to be brought into agreement with observation, force the number of components to be raised from two to four. We shall see that two components suffice if the requirement of left-right symmetry (parity) is dropped.

Two-component Theory

§1. *Transformation Law for* ψ. If one introduces the projective coordinates x_α into a space with Cartesian coordinates x, y, z:

$$x = \frac{x_1}{x_o}, \qquad y = \frac{x_2}{x_o}, \qquad z = \frac{x_3}{x_o},$$

the equation for the unit sphere becomes

$$-x_o^2 + x_1^2 + x_2^2 + x_3^2 = 0. \tag{1}$$

If one projects from the south pole to the equatorial plane $z = 0$ which is regarded as the carrier of the complex variables

$$x + iy = \zeta = \frac{\psi_2}{\psi_1}$$

the equations may be written as

$$\begin{aligned} x_o &= \bar{\psi}_1\psi_1 + \bar{\psi}_2\psi_2, & x_1 &= \bar{\psi}_1\psi_2 + \bar{\psi}_2\psi_1, \\ x_2 &= i(-\bar{\psi}_1\psi_2 + \bar{\psi}_2\psi_1), & x_3 &= i(\bar{\psi}_1\psi_1 - \bar{\psi}_2\psi_2). \end{aligned} \tag{2}$$

Thus the x_α are hermitian forms of ψ_1, ψ_2. Here, only the ratios of the variables ψ_1, ψ_2 and the coordinates x_α are relevant. A homogeneous linear transformation of ψ_1, ψ_2 (with complex coefficients) induces a real linear transformation among the coordinates x_α: it represents a collineation that leaves the unit sphere and the sense of rotation on the unit sphere invariant. It is well-known and easy to show that every such collineation can be obtained uniquely in this way.

Going from the homogeneous to the inhomogeneous, one now considers the x_α as coordinates in a four-dimensional world and (1) as the equation for the "light-cone"; and one restricts oneself to those linear transformations U of ψ_1, ψ_2 for which the determinant has absolute value 1. U induces a Lorentz transformation on the x_α, that is a real homogeneous transformation, which leaves the form

$$-x_o^2 + x_1^2 + x_2^2 + x_3^2$$

invariant. However, the formula for x_0 and our remarks about the invariance of the sense of rotation tell us immediately that the Lorentz transformations contain only a single closed continuous set Λ which (1) preserve past and future and (2) have determinant 1 and not -1; but are otherwise unconstrained. The linear transformation U is not fully determined by Λ, an arbitrary factor $e^{i\lambda}$ of absolute value 1 being left at our disposal. One may normalize it by the requirement that the determinant of U is equal to 1 but even then it remains double-valued. One would like to preserve restriction 1; it is one of the most hopeful features of the ψ-theory that it can take account of the difference between past and future. The restriction 2 removes the equivalence of left and right. It is only the fact that left-right symmetry actually appears in Nature that forces us (Part II) to introduce a second pair of ψ-components.

The hermitian conjugate of a matrix $A = \|a_{ik}\|$ will be denoted by A^*:

$$a_{ik}^* = \bar{a}_{ki}.$$

Let S_α denote the coefficient-matrix of the hermitian form of the variables ψ_1, ψ_2 which represents the coordinates x_α in (2):

$$x_\alpha = \psi^* S_\alpha \psi; \tag{3}$$

where ψ denotes the column-vector ψ_1, ψ_2. Let S_0 denote the unit matrix; we have the equations

$$S_1^2 = 1, \qquad S_2 S_3 = i S_1 \tag{4}$$

and those obtained from them by cyclic permutation of the indices. From the formal point of view it is more convenient to replace the real time-component x_0 by the imaginary ix_0. The Lorentz transformations then appear as orthogonal transformations of the four quantities

$$x(o) = ix_o, \qquad x(\alpha) = x_\alpha \qquad [\alpha = 1, 2, 3].$$

Instead of (3) one writes

$$x(\alpha) = \psi^* S(\alpha)\psi. \tag{5}$$

The transformation law of the ψ-components is such that under the influence of a transformation Λ of the coordinates $x(\alpha)$ they change so that the quantities (5) experience the transformation Λ. A quantity of this kind represents the wave-function of a matter-field, as we know from the spin-phenomenon. The $x(\alpha)$ are the coordinates in an "orthogonal tetrad" $\mathbf{e}(\alpha)$; $\mathbf{e}(1), \mathbf{e}(2), \mathbf{e}(3)$ are real space–like vectors, which constitute a left-handed Cartesian coordinate-system, $\mathbf{e}(0)/i$ is a real time–like vector in the future direction. The transformation Λ describes the transformation from such a coordinate system to another of the same kind, which will be called from now on a rotation of the tetrad. We obtain the same coefficients $c(\alpha, \beta)$ whether we express the transformation Λ through the base-vectors of the tetrad or through the coordinates

$$\mathbf{x} = \sum_\alpha x(\alpha)\mathbf{e}(\alpha) = \sum_\alpha x'(\alpha)\mathbf{e}'(\alpha)$$
$$\mathbf{e}'(\alpha) = \sum_\beta c(\alpha\beta)\mathbf{e}(\beta), \qquad x'(\alpha) = \sum_\beta c(\alpha\beta)x(\beta);$$

This follows from the orthogonal character of Λ.

For the following it is necessary to compute the infinitesimal transformation

$$d\psi = dE.\psi \tag{6}$$

that corresponds to an arbitrary infinitesimal rotation $d\Omega$:

$$dx(\alpha) = \sum_\beta do(\alpha\beta).x(\beta).$$

The $do(\alpha, \beta)$ form an anti-symmetric matrix. The transformation (6) is normalized so that the trace of dE is 0. The matrix dE depends linearly and homogeneously on the $do(\alpha, \beta)$; we therefore write

$$dE = \frac{1}{2}\sum_{\alpha\beta} do(\alpha\beta).A(\alpha\beta) = \sum do(\alpha\beta).A(\alpha\beta).$$

The last sum is to be taken only over the pairs

$$(\alpha\beta) = (01), (02), (03); \quad (23), (31), (12).$$

The dependence of $A(\alpha, \beta)$ on α and β is, of course, anti-symmetric. It should not be forgotten that the coefficients $do(\alpha, \beta)$ are pure imaginary for the first three pairs and real for the second three pairs, but are otherwise arbitrary. One finds that

$$A(23) = -\frac{1}{2i}S(1), \qquad A(01) = \frac{1}{2i}S(1) \tag{7}$$

together with two analogous equations that are obtained by cyclic permutations of the indices. To establish this one simply has to check that the infinitesimal transformations dE

$$d\psi = \frac{1}{2i}S(1)\psi \quad \text{and} \quad d\psi = \frac{1}{2}S(1)\psi$$

generate the infinitesimal rotations

$$dx(0) = 0, \quad dx(1) = 0, \quad dx(2) = -x(3), \quad dx(3) = x(2)$$

and

$$dx(0) = ix(1), \quad dx(1) = -ix(0), \quad dx(2) = 0, \quad dx(3) = 0$$

respectively.

§2. *Metric and Parallel Displacement.* We now consider General Relativity. We describe the metric at a point P by giving a local normal tetrad $\mathbf{e}(\alpha)$. Only the class of local normal tetrads—which are connected to each other by the group of rotations Λ—is determined by the metric; an individual member of the class is selected arbitrarily. The laws are invariant with respect to an arbitrary rotation of the local tetrad; the rotation of the tetrad in another point P' being independent of the rotation at P. Let $\psi_1(P)$, $\psi_2(P)$ be the components of the matter-potential at the point P relative to the chosen tetrad. A vector $\mathbf{t}$ at P can be written in the form

$$\mathbf{t} = \sum_\alpha t(\alpha)\mathbf{e}(\alpha);$$

the numbers $t(\alpha)$ being its components relative to the tetrad. For an analytic representation we need in addition a coordinate system x_p; x_p are any four continuous functions of position whose values serve to distinguish different points. The laws are therefore invariant with respect to arbitrary coordinate transformations. Let $e^p(\alpha)$ be the components of $\mathbf{e}(\alpha)$ in the coordinate-system. These four 4-component quantities describe the gravitational field. The contravariant components t^p of a vector $\mathbf{t}$ with respect to the coordinate system are connected to its components $t(\alpha)$ relative to the tetrad by the equation

$$t^p = \sum_\alpha t(\alpha)e^p(\alpha);$$

Conversely, the $t(\alpha)$ are computed from the covariant components t_p in the coordinate system according to

$$t(\alpha) = \sum_p t_p e^p(\alpha).$$

These equations control the interchange of indices. I have written the Greek indices which refer to the tetrad as arguments because for them there is no distinction between upper and lower. The interchange in the opposite sense proceeds through the matrix $\|e_p(\alpha)\|$ inverse to $\|e^p(\alpha)\|$:

$$\sum_\alpha e_p(\alpha)e^q(\alpha) = \delta_p^q \quad \text{and} \quad \sum_p e_p(\alpha)e^p(\beta) = \delta(\alpha, \beta).$$

where δ is either 0 or 1 according [to whether] the indices are the same or not. The summation rule will be used for both Latin and Greek indices. Let ϵ denote the absolute value of the determinant $|e^p(\alpha)|$. A quantity denoted by a Latin letter and divided by ϵ will, as usual, be denoted by the corresponding German letter; e.g.

$$\mathfrak{e}^p(\alpha) = \frac{e^p(\alpha)}{\epsilon}.$$

Vectors and tensors may be described by either their coordinate or tetrad components. For the quantity ψ, however, only the tetrad components make sense because the transformation of the components is according to a representation of the rotation group that cannot be extended to the group of all linear transformations. This is why, in the theory of matter, it is necessary [2] to represent the gravitational field analytically in the above form instead of the metrical form

$$\sum_{p,q} g_{pq} dx_p dx_q.$$

In addition

$$g_{pq} = e_p(\alpha)e_q(\alpha).$$

Gravitational theory must be re-cast in this new analytic form. I begin with the formula for the infinitesimal parallel transport determined by the metric. Let the vector $\mathbf{e}(\alpha)$ at the point P be transported into the vector $\mathbf{e}'(\alpha)$ at the infinitely close point P'. The $\mathbf{e}'(\alpha)$ form a normal tetrad at P' which is obtained from the local tetrad $\mathbf{e}(\alpha) = \mathbf{e}(\alpha; P')$ by an infinitesimal rotation $d\Omega$

$$\delta\mathbf{e}(\beta) = \sum_\gamma do(\beta\gamma).\mathbf{e}(\gamma), \qquad \delta\mathbf{e}(\beta) = \mathbf{e}'(\beta) - \mathbf{e}(\beta; P'). \tag{8}$$

$d\boldsymbol{\Omega}$ is linearly dependent on the transfer PP' or its components,

$$dx_p = (dx)^p = v^p = e^p(\alpha)v(\alpha).$$

Hence we write

$$d\boldsymbol{\Omega} = \boldsymbol{\Omega}_p(dx)^p, \quad do(\beta\gamma) = o_p(\beta\gamma)(dx)^p = o(\alpha;\beta\gamma)v(\alpha). \tag{9}$$

As is well known the parallel transfer of the vector $\mathbf{t}$ with components t^p is described by an equation

$$d\mathbf{t} = -d\boldsymbol{\Gamma}.\mathbf{t}, \quad \text{i.e.} \quad dt^p = -d\Gamma^p_r.t^r, \quad d\Gamma^p_r = \Gamma^p_{rq}(dx)^q,$$

in which the symbols Γ^p_{rq}, which are independent of $\mathbf{t}$ and dx, are symmetric in r and q. We therefore have

$$\mathbf{e}'(\beta) - \mathbf{e}(\beta) = -d\boldsymbol{\Gamma}.\mathbf{e}(\beta)$$

in addition to equation (8). Subtraction of the two differences on the left-hand-side produces the differential $d\mathbf{e}(\beta) = \mathbf{e}(\beta;\, P') - \mathbf{e}(\beta;\, P)$:

$$d\mathbf{e}^p(\beta) + d\boldsymbol{\Gamma}^p_r e^r(\beta) = -do(\beta\gamma).e^p(\gamma),$$
$$\frac{\partial e^p(\beta)}{\partial x_q}.e^q(\alpha) + \Gamma^p_{rq}e^r(\beta)e^q(\alpha) = -o(\alpha;\beta\gamma)e^p(\gamma).$$

Here one may eliminate the o to obtain the known equation for determining the $\boldsymbol{\Gamma}$ if one expresses the fact that $o(\alpha;\beta\gamma)$ is anti-symmetric with respect to β and γ. One eliminates the $\boldsymbol{\Gamma}$ and computes o by making use of the fact that Γ^p_{rq} is symmetric with respect to r and q i.e. that

$$\Gamma^p(\beta,\alpha) = \Gamma^p_{rq}e^r(\beta)e^q(\alpha)$$

is symmetric in α and β:

$$\frac{\partial e^p(\alpha)}{\partial x_q}e^q(\beta) - \frac{\partial e^p(\beta)}{\partial x_q}e^q(\alpha) = \{o(\alpha;\beta\gamma) - o(\beta;\alpha\gamma)\}e^p(\gamma). \tag{10}$$

The left-hand side consists of the components of that "commutator-product" of the vector-fields $\mathbf{e}(\alpha), \mathbf{e}(\beta)$ which is invariant with respect to coordinate transformations and plays the decisive role in the Lie theory of infinitesimal transformations; it will be denoted by $[\mathbf{e}(\alpha), \mathbf{e}(\beta)]$. Since $o(\beta; \alpha\gamma)$ is anti-symmetric in α and γ one has

$$[\mathbf{e}(\alpha), \mathbf{e}(\beta)]^p = \{o(\alpha; \beta\gamma) + o(\beta; \gamma\alpha)\}e^p(\gamma)$$

or

$$o(\alpha; \beta\gamma) + o(\beta; \gamma\alpha) = [\mathbf{e}(\alpha), \mathbf{e}(\beta)](\gamma). \tag{11}$$

If one takes the three cyclic permutations of $\alpha\beta\gamma$ in this equation and adds the resultant equations with the signs $+ - +$ one obtains

$$2o(\alpha; \beta\gamma) = [\mathbf{e}(\alpha), \mathbf{e}(\beta)](\gamma) - [\mathbf{e}(\beta), \mathbf{e}(\gamma)](\alpha) + [\mathbf{e}(\gamma), \mathbf{e}(\alpha)](\beta),$$

which shows that $o(\alpha : \beta\gamma)$ is, in fact, uniquely determined. The expression obtained fulfils all the conditions, since, as easily seen, it is anti-symmetric in β and γ.

For the following we shall need in particular the contractions

$$o(\rho, \rho\alpha) = [\mathbf{e}(\alpha), \mathbf{e}(\rho)](\rho) = \frac{\partial e^p(\alpha)}{\partial x_p} - \frac{\partial e^p(\rho)}{\partial x_q} e^q(\alpha) e_p(\rho).$$

Since

$$-\epsilon.\delta\left(\frac{1}{\epsilon}\right) = \frac{\delta\epsilon}{\epsilon} = e_q(\rho).\delta e^q(\rho)$$

we have

$$o(\rho, \rho\alpha) = \epsilon \frac{\partial e^p(\alpha)}{\partial x^p}. \tag{12}$$

§3. *Action of Matter.* With the help of parallel-transport we can compute the covariant derivative not only of the vector and tensor fields but also the ψ-field. Let $\psi_a(P)$, $\psi_a(P')$ $[a = 1, 2]$ be the components relative to the local tetrad $\mathbf{e}(\alpha)$ at P and P' respectively. The difference $\psi_a(P') - \psi_a(P) = d\psi_a$ is the usual differential. On the other hand we parallel-transport the tetrad $\mathbf{e}(\alpha)$ from P to P': $\mathbf{e}'(\alpha)$; and let ψ'_α be the components of ψ at P' with respect to the tetrad $\mathbf{e}'(\alpha)$ itself. ψ_α and ψ'_α depend only on the choice of the tetrad $\mathbf{e}(\alpha)$ at P; they have nothing to do with the local tetrad at P'. On rotation of the tetrad at P the ψ'_α transform in exactly the same way as the ψ_α , and similarly for the differentials $\delta\psi_\alpha = \psi'_\alpha - \psi_\alpha$. They are the components of the covariant differential $\delta\psi$ of ψ. $\mathbf{e}'(\alpha)$ is obtained from the local tetrad $\mathbf{e}(\alpha) = \mathbf{e}(\alpha; P')$ at P' by the infinitesimal rotation determined in §2. The corresponding infinitesimal transformation

$$dE = \frac{1}{2} do(\beta\gamma).A(\beta\gamma)$$

changes $\psi_\alpha(P')$ into ψ'_α, i.e. $\psi' - \psi(P)$ is equal to $dE.\psi$. If one adds $d\psi = \psi(P') - \psi(P)$ one obtains

$$\delta\psi = d\psi + dE.\psi. \tag{13}$$

Everything depends linearly on the displacement PP'. We shall write

$$\delta\psi = \psi_p(dx)^p = \psi(\alpha)v(\alpha), \qquad dE = E_p(dx)^p = E(\alpha)v(\alpha).$$

We find

$$\psi_p = \left(\frac{\partial}{\partial x_p} + E_p\right)\psi \quad \text{or} \quad \psi(\alpha) = \left(e^p(\alpha)\frac{\partial}{\partial x_p} + E(\alpha)\right)\psi,$$

where

$$E(\alpha) = \frac{1}{2}o(\alpha;\beta\gamma)A(\beta\gamma).$$

If ψ' is a quantity with the same transformation law as ψ the

$$\psi^* S(\alpha)\psi'$$

are the components of a vector with respect to the local tetrad. Hence

$$v'(\alpha) = \psi^* S(\alpha)\delta\psi = \psi^* S(\alpha)\psi(\beta).v(\beta)$$

is a tetrad-independent linear transformation $v \to v'$ of the vector space at P. Its trace

$$\psi^* S(\alpha)\psi(\alpha)$$

is therefore a scalar and the equation

$$i\epsilon\mathfrak{m} = \psi^* S(\alpha)\psi(\alpha) \tag{14}$$

defines a scalar density $\mathfrak{m}$ whose integral

$$\int \mathfrak{m}dx \qquad\qquad (dx = dx_0dx_1dx_2dx_3)$$

can be used as an action. To obtain an explicit expression for $\mathfrak{m}$ we must compute

$$S(\alpha)E(\alpha) = \frac{1}{2}S(\alpha)A(\beta\gamma)o(\alpha;\beta\gamma). \tag{15}$$

From (7) and (4) it follows that

$$S(\beta)A(\beta\alpha) = \frac{1}{2}S(\alpha) \qquad [\alpha \neq \beta, \quad \text{no summation over} \quad \beta]$$

and

$$S(\beta)A(\gamma\delta) = \frac{1}{2}S(\alpha),$$

if $\alpha\beta\gamma\delta$ is an even permutation of the indices. The terms of the first and second kind produce as a contribution to (15) the following multiples of $S(\alpha)$

$$\frac{1}{2}o(\rho;\rho\alpha) = \frac{1}{2\epsilon}\frac{\partial \mathfrak{e}^p(\alpha)}{\partial x_p}$$

and

$$o(\beta;\gamma\delta) + o(\gamma;\delta\beta) + o(\delta;\beta\gamma) = \frac{i}{2}\phi(\alpha)$$

respectively. When $\alpha\beta\gamma\delta$ is an even permutation of 0123 we have from (11)

$$\begin{aligned} i\phi(\alpha) &= [\mathbf{e}(\beta), \mathbf{e}(\gamma)](\delta) + (\text{cycl perm of } \beta\gamma\delta) \\ &= \sum + \frac{\partial e^p(\beta)}{\partial x_q} e^q(\gamma) e_p(\delta). \end{aligned} \tag{16}$$

The sum extends over the six permutations of $\beta\gamma\delta$ with alternating sign (as well as p and q). With this notation we have

$$\mathfrak{m} = \frac{1}{i}\left(\psi^*\mathfrak{e}^p(\alpha)S(\alpha)\frac{\partial\psi}{\partial x_p} + \frac{1}{2}\frac{\partial\mathfrak{e}^p(\alpha)}{\partial x_p}\psi^* S(\alpha)\psi\right) + \frac{1}{4\epsilon}\phi(\alpha)s(\alpha). \tag{17}$$

The second part is

$$= \frac{1}{4i\epsilon}\left|e_p(\alpha), e^q(\alpha), \frac{\partial e^p(\alpha)}{\partial x_q}, s(\alpha)\right|$$

(summed over p and q); each term is a determinant of four lines, which one obtains from the given line by setting $\alpha = 0, 1, 2, 3$ successively,

$$s(\alpha) \quad \text{is} \quad = \psi^* S(\alpha)\psi \tag{18}$$

It is not the action-integral

$$\int \mathfrak{h}\,dx \tag{19}$$

itself but only its variation that is important for the equations of motion. Hence it is not necessary that $\mathfrak{h}$ be real but only that the difference $\bar{\mathfrak{h}} - \mathfrak{h}$ be a divergence. In that case we shall say that $\mathfrak{h}$ is effectively real. We must therefore check how $\mathfrak{m}$ behaves from this point of view. $e^p(\alpha)$ is real for $\alpha = 1, 2, 3$, pure imaginary for $\alpha = 0$. Hence $e^p(\alpha)S(\alpha)$ is a hermitian matrix. Similarly $\phi(\alpha)$ is real for $\alpha = 1, 2, 3$, pure imaginary for $\alpha = 0$; Hence $\phi(\alpha)S(\alpha)$ is also hermitian. Accordingly

$$\bar{\mathfrak{m}} = -\frac{1}{i}\left(\frac{\partial \psi^*}{\partial x_p}\mathfrak{S}^p\psi + \frac{1}{2}\frac{\partial \mathfrak{e}^p(\alpha)}{\partial x_p}\psi^* S(\alpha)\psi\right) + \frac{1}{4\epsilon}\phi(\alpha)s(\alpha),$$

$$i(\mathfrak{m} - \bar{\mathfrak{m}}) = \psi^*\mathfrak{S}^p\frac{\partial \psi}{\partial x_p} + \frac{\partial \psi^*}{\partial x_p}\mathfrak{S}^p\psi + \frac{\partial \mathfrak{e}^p(\alpha)}{\partial x_p}\psi^* S(\alpha)\psi$$

$$= \frac{\partial}{\partial x_p}(\psi^*\mathfrak{S}^p\psi) = \frac{\partial \mathfrak{s}^p}{\partial x_p}$$

Thus $\mathfrak{m}$ is in fact effectively real. We recover the special relativistic case when we set

$$e^o(0) = -i, \quad e^1(1) = e^2(2) = e^3(3) = 1,$$

with all other $e^p(\alpha) = 0$.

§4. *Energy*. Let (19) be the action-integral for the matter in the broad sense (matter and electric field) which are to be described by ψ and the electromagnetic potential f_p. The equations of motion are obtained from the requirement that the variation be zero

$$\delta \int \mathfrak{h} dx = 0$$

for arbitrary infinitesimal variations of ψ and f_p which vanish outside a finite space-time volume. The variation of ψ gives the field-equations for the matter in the narrower sense, the variation of the f_p the field-equations for the electromagnetic field. On subjecting the $e^p(\alpha)$, which up to now were kept invariant, to an analogous variation, one obtains, on account of these field equations, an equation

$$\delta \int \mathfrak{h} dx = \int \mathfrak{t}_p(\alpha).\delta e^p(\alpha).dx, \tag{20}$$

which defines the energy tensor density $\mathfrak{t}_p(\alpha)$.

Because of the invariance of the action the expression (20) must vanish when the variations $\delta e^p(\alpha)$ are produced by

(1) the local tetrad $\mathbf{e}(\alpha)$ undergoing an infinitesimal rotation while the coordinate-system x_p is kept fixed; or

(2) the coordinates x_p undergoing an infinitesimal transformation while the tetrad is kept fixed.

The first procedure is described by the equation

$$\delta e^p(\alpha) = o(\alpha\beta).e^p(\beta).$$

Here the $o(\alpha\beta)$ form an anti-symmetric (infinitesimal) matrix whose dependence on position is arbitrary. And the vanishing of (20) implies that

$$\mathfrak{t}(\beta\alpha) = \mathfrak{t}_p(\alpha)e^p(\beta)$$

is symmetric in α and β. The symmetry of the energy tensor is thus equivalent to the first invariance property. The symmetry is not an identity but is a consequence of the electromagnetic and matter field-equations, since for a fixed ψ-field the components will change under a rotation of the tetrad!

It is somewhat more tedious to compute the $\delta e^p(\alpha)$ that is produced by the second variation. But the computations are similar to those used in the standard analytical formulation of relativity theory [3]. Let the transformed coordinates of the point P be

$$x'_p = x_p + \delta x_p, \qquad \delta x_p = \xi^p(x).$$

Let P' denote the point that in the new coordinate-system has the same coordinates x_p that P had in the old one; it has in the old system the coordinates $x_p - \delta x_p$. The vector $\mathbf{t}$ at P will have the components

$$\frac{\partial x'_p}{\partial x_q} t^q = t_p + \frac{\partial \xi^p}{\partial x_q} t^q$$

in the new coordinate system. In particular the changes experienced by the components $e^p(\alpha)$ of the fixed vector $\mathbf{e}(\alpha)$ at the fixed point P are

$$\delta' e^p(\alpha) = \frac{\partial \xi^p}{\partial x_q} e^q(\alpha)$$

On the other hand the difference between the vector $\mathbf{e}(\alpha)$ at P' and P is given by

$$de^p(\alpha) = -\frac{\partial e^p}{\partial x_q} \xi^q.$$

Hence the variation produced by the coordinate-transformation for fixed coordinate values x_p is:

$$\delta e^p(\alpha) = \frac{\partial \xi^p}{\partial x_q} e^q(\alpha) - \frac{\partial e^p(\alpha)}{\partial x_q} \xi^q.$$

Here the ξ^p are arbitrary, apart from a function that vanishes outside a finite region of space-time. If we substitute in (20) we obtain by a partial integration

$$0 = \int \left\{ \frac{\partial \mathfrak{t}_p^q}{\partial x_q} + \mathfrak{t}_q(\alpha) \frac{\partial e^q(\alpha)}{\partial x_p} \right\} \xi^p dx.$$

The quasi-conservation law of energy and momentum follows from this in the form

$$\frac{\partial \mathfrak{t}_p^q}{\partial x_q} + \frac{\partial e^q(\alpha)}{\partial x_p} \mathfrak{t}_q(\alpha) = 0. \tag{21}$$

On account of the second term this is a true conservation equation only in the case of special relativity. In general relativity it will be so only if the energy of the gravitational field is added. In special relativity integration over the space-like surface

$$x_o = t = \text{const} \tag{22}$$

with measure $d\xi = dx_1 dx_2 dx_3$ yields the time-independent components of the momentum (J_1, J_2, J_3) and energy $(-J_0)$:

$$J_p = \int \mathfrak{t}_p^o d\xi.$$

With the help of the symmetry one finds also the divergence equations

$$\frac{\partial}{\partial x_q}(x_2 \mathfrak{t}_3^q - x_3 \mathfrak{t}_2^q) = 0, \ldots,$$
$$\frac{\partial}{\partial x_q}(x_o \mathfrak{t}_1^q + x_1 \mathfrak{t}_o^q) = 0, \ldots,$$

The three equations of the first kind show that the angular momentum (M_1, M_2, M_3)

$$M_1 = \int (x_2 \mathfrak{t}_3^o - x_3 \mathfrak{t}_2^o) d\xi, \ldots,$$

is constant in time, the equations of the second kind contain the theorem of the inertia of energy.

We compute the energy density for the matter-action $\mathfrak{m}$ constructed above; we handle the two parts which appear in the decomposition of $\mathfrak{m}$ in (17) separately. For the first part we obtain after partial integration

$$\int \delta \mathfrak{m}.dx = \int u_p(\alpha) \delta \mathfrak{e}^p(\alpha).dx$$

with

$$iu_p(\alpha) = \psi^* S(\alpha)\frac{\partial\psi}{\partial x_p} - \frac{1}{2}\frac{\partial(\psi^* S(\alpha)\psi)}{\partial x_p},$$
$$= \frac{1}{2}\left(\psi^* S(\alpha)\frac{\partial\psi}{\partial x_p} - \frac{\partial\psi^*}{\partial x_p}S(\alpha)\psi\right).$$

The part of the energy coming from this term is therefore

$$\mathfrak{t}_p(\alpha) = \mathfrak{u}_p(\alpha) - e_p(\alpha).\mathfrak{u}, \qquad \mathfrak{t}_p^q = \mathfrak{u}_p^q - \delta_p^q \mathfrak{u},$$

where $\mathfrak{u}$ is the contraction $e^p(\alpha)\mathfrak{u}_p(\alpha)$. These formulae are correct also for non-constant $e^p(\alpha)$. In the second part, however, we restrict ourselves for simplicity to special relativity. In this case

$$\int \delta\mathfrak{m}.dx = \frac{1}{4i}\int \left| e_p(\alpha), e^q(\alpha), \frac{\partial\,(\delta e^p(\alpha))}{\partial x_q}, s(\alpha) \right| dx$$
$$= -\frac{1}{4i}\int \left| \delta e^q(\alpha), e_p(\alpha), e^q(\alpha), \frac{\partial s(\alpha)}{\partial x_q} \right| dx$$
$$\mathfrak{t}_p(0) = -\frac{1}{4i}\left| e_p(\alpha), e^q(\alpha), \frac{\partial s(\alpha)}{\partial x_q} \right|_{\alpha=1,2,3}.$$

From this we obtain $\mathfrak{t}_p^0$ by multiplication with $-i$; thus $\mathfrak{t}_0^0 = 0$ and

$$\mathfrak{t}_1^o = \frac{1}{4}\left(\frac{\partial s(3)}{\partial x_2} - \frac{\partial s(2)}{\partial x_3}\right). \tag{23}$$

We combine the two parts to determine the total energy, momentum and angular momentum. From

$$\mathfrak{t}_o^o = -\frac{1}{2i}\sum_{p=1}^{3}\left(\psi^* S^p \frac{\partial\psi}{\partial x_p} - \frac{\partial\psi^*}{\partial x_p}S^p\psi\right)$$

one obtains after carrying out a partial integration on the subtracted term,

$$-J_o = -\int \mathfrak{t}_o^o d\xi = \frac{1}{i}\int \psi^* \sum_{p=1}^{3} S^p \frac{\partial\psi}{\partial x_p} d\xi$$

This leads us to consider the operator

$$\frac{1}{i}\sum_{p=1}^{3} S^p \frac{\partial}{\partial x_p}$$

as the representative of the energy of a free particle. We also have

$$J_1 = \int \mathfrak{t}_1^o d\xi = \frac{1}{2i} \int \left(\psi^* \frac{\partial \psi}{\partial x_1} - \frac{\partial \psi^*}{\partial x_1} \psi \right) d\xi$$
$$= \frac{1}{i} \int \psi^* \frac{\partial \psi}{\partial x_1} d\xi.$$

The term (23) does not contribute to the integral. The momentum is represented by the operator

$$\frac{1}{i} \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right)$$

as it should be according to Schrödinger. From the full expression for

$$x_2 \mathfrak{t}_3^o - x_3 \mathfrak{t}_2^o$$

one obtains finally by suitable partial integrations

$$M_1 = \int \left\{ \frac{1}{i} \psi^* \left(x_2 \frac{\partial \psi}{\partial x_3} - x_3 \frac{\partial \psi}{\partial x_2} \right) + \frac{1}{2} s(1) \right\} . d\xi.$$

In agreement with known formulae M_1 is therefore represented by the operator

$$\frac{1}{i} \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right) + \frac{1}{2} S(1)$$

Since one has incorporated spin in the theory from the beginning it must reappear here; it is, however, surprising and instructive how that happens. One sees that the fundamental assumptions of quantum-mechanics are based less on principle than one had originally supposed. They are connected to the special action $\mathfrak{m}$. On the other hand this connection confirms the necessity of $\mathfrak{m}$ in its role of matter-action. Only general relativity, which leads to a unique definition of the energy through the free variation of $e^p(\alpha)$, permits us to close the circle of quantum-mechanics as was shown above.

5. *Gravitation.* We return to the transcription of Einstein's classical theory of gravitation and determine first the curvature tensor [4]. Let the line elements d and δ connect the point P to the points P_d and P_δ. The line elements δ and d respectively are then transferred to P_d and P_δ along arbitrary paths so that they meet at the corner P^* opposite to P of an infinitesimal 'parallelogram'. Let the tetrad $\mathbf{e}(\alpha)$ be parallel-transferred along the paths PP_dP^* and $PP_\delta P^*$

respectively. The two normal tetrads that one obtains in this way at P^* can be obtained from another by an infinitesimal rotation

$$\mathbf{P}_{pq}(dx)^p(\delta x)^q = \frac{1}{2}\mathbf{P}_{pq}(\Delta x)^{pq}$$

where

$$(\Delta x)^{pq} = (dx)^p(\delta x)^q - (\delta x)^p(dx)^q$$

are the components of the surface spanned by dx and δx and P_{pq} is anti-symmetric with respect to p and q. P_{pq} is an anti-symmetric matrix $\|r_{pq}(\alpha\beta)\|$; it is the Riemann curvature tensor.

The rotation that produces the tetrad $\mathbf{e}^*(\alpha)$ from the local tetrad $\mathbf{e}(\alpha)$ after parallel-transfer to P^* along the first path is, in easily understood notation,

$$(1 + d\boldsymbol{\Omega})(1 + \delta\boldsymbol{\Omega}(P_d)).$$

The difference between this expression and the one obtained by interchange of d and δ is

$$= (d(\delta\boldsymbol{\Omega}) - \delta(d\boldsymbol{\Omega})) + (d\boldsymbol{\Omega}\delta\boldsymbol{\Omega} - \delta\boldsymbol{\Omega}d\boldsymbol{\Omega}).$$

$$d\boldsymbol{\Omega} = \boldsymbol{\Omega}_p(dx)^p,$$

$$\delta(d\boldsymbol{\Omega}) = \frac{\partial\boldsymbol{\Omega}_p}{\partial x_q}\delta x_q dx_p + \boldsymbol{\Omega}_p \delta dx_p.$$

Since the parallelogram closes we have $\delta dx_p = d\delta x_p$; Hence finally

$$\mathbf{P}_{pq} = \left(\frac{\partial\boldsymbol{\Omega}_p}{\partial x_q} - \frac{\partial\boldsymbol{\Omega}_q}{\partial x_p}\right) + \left(\boldsymbol{\Omega}_p\boldsymbol{\Omega}_q - \boldsymbol{\Omega}_q\boldsymbol{\Omega}_p\right).$$

For the scalar curvature

$$r = e^p(\alpha)e^q(\beta)r_{pq}(\alpha\beta)$$

the first, differentiated, part gives the contribution

$$\left(e^q(\alpha)e^q(\beta) - e^q(\beta)e^p(\alpha)\right)\frac{\partial o_p(\alpha\beta)}{\partial x_q}.$$

For $\mathfrak{r} = \frac{r}{\epsilon}$ it gives, after dropping a total divergence, the two terms

$$-2o(\beta;\alpha\beta)\frac{\partial \mathfrak{e}^p(\beta)}{\partial x_q},$$

and

$$\frac{1}{\epsilon} o_p(\alpha\beta) \left\{ \frac{\partial e^p(\alpha)}{\partial x_q} e^q(\beta) - \frac{\partial e^p(\beta)}{\partial x_q} e^q(\alpha) \right\}.$$

The first is, according to (12)

$$= -2o(\beta;\, \rho\beta) o(\alpha;\, \alpha\rho).$$

The second, according to (10)

$$= 2o(\alpha;\, \beta\gamma).o(\gamma;\, \alpha\beta).$$

The result is the following expression for the gravitational action-density $\mathfrak{g}$

$$\epsilon \mathfrak{g} = o(\alpha;\, \beta\gamma) o(\gamma;\, \alpha\beta) + o(\alpha;\, \alpha\gamma) o(\beta;\, \beta\gamma). \tag{24}$$

The integral $\int \mathfrak{g} dx$ is not truly invariant, but is effectively invariant as $\mathfrak{g}$ differs from the scalar density $\mathfrak{r}$ by a divergence.

Variation of the total action integral

$$\int (\mathfrak{g} + \kappa \mathfrak{h}) dx$$

yields the gravitational equations (κ being a numerical constant).

One obtains from $\mathfrak{g}$ the gravitational energy $\mathfrak{v}_p^q$ by considering an infinitesimal displacement in coordinate space [5]

$$x'_p = x_p + \xi^p, \qquad \xi^p = \text{const.}$$

The resultant variation

$$\delta \mathbf{e}(\alpha) \quad \text{is} \quad = -\frac{\partial \mathbf{e}(\alpha)}{\partial x_p} \xi^p.$$

$\mathfrak{g}$ is a function of $e^p(\alpha)$ and the derivatives $e_q^p(\alpha) = \frac{\partial e^p(\alpha)}{\partial x_q}$; the total derivative will be denoted by

$$\delta \mathfrak{g} = \mathfrak{g}_p(\alpha) \delta e^p(\alpha) + \mathfrak{g}_p^q(\alpha) \delta e_q^p(\alpha).$$

For the variations due to the infinitesimal translations in coordinate space we must have

$$\int \delta \mathfrak{g} dx + \int \frac{\partial \mathfrak{g}}{\partial x_p} \xi^p dx = 0, \tag{25}$$

the integral being taken over an arbitrary volume of space-time,

$$\int \delta \mathfrak{g} dx = \int \left(\mathfrak{g}(\alpha) - \frac{\partial \mathfrak{g}_p^q(\alpha)}{\partial x_q} \right) \delta e^p(\alpha) dx + \int \frac{\partial \left(\mathfrak{g}_p^q(\alpha) \delta e^p(\alpha) \right)}{\partial x_q} dx.$$

According to the gravitational equations the bracket in the first integral $=$ $-\kappa \mathfrak{t}_p(\alpha)$, the integral itself

$$= -\kappa \int \mathfrak{t}_q(\alpha) \frac{\partial e^q(\alpha)}{\partial x_q} \xi^p dx.$$

One introduces

$$\mathfrak{v}_p^q = \delta_p^q \mathfrak{g} - \frac{\partial e^r(\alpha)}{\partial x_p} \mathfrak{g}_r^q(\alpha).$$

Equation (25) shows that the integral of

$$\left(\mathfrak{v}_p^q - \kappa \mathfrak{t}_q(\alpha) \frac{\partial e^q(\alpha)}{\partial x_p} \right) \xi^p$$

taken over an arbitrary volume of space-time is zero. The integrand must therefore vanish everywhere. Since the ξ^p are arbitrary constants the coefficients of the ξ^p are separately zero. In this way (21) is converted into a pure divergence equation

$$\frac{\partial \left(\mathfrak{v}_p^q + \kappa \mathfrak{t}_p^q \right)}{\partial x_q} = 0$$

and $\mathfrak{v}_q^p/\kappa$ is seen to be the gravitational energy. To formulate in addition a true differential conservation law for the angular momentum in general relativity one must choose coordinates such that the contragredient rotation of all tetrads appears as an orthogonal transformation of the coordinates. This is certainly possible, but I shall not pursue this question further here.

6. *Electric Field.* We come now to the critical part of the theory. In my opinion the origin and necessity for the electromagnetic field is the following. The components ψ_1, ψ_2 are, in fact, not uniquely determined by the tetrad but only to the extent that they can still be multiplied by an arbitrary "gauge-factor" $e^{i\lambda}$. The transformation of the ψ induced by a rotation of the tetrad is determined only up to such a factor. In special relativity one must regard this gauge-factor as a constant because here we have only a single point-independent tetrad. Not so in general relativity; every point has its own tetrad and hence its own arbitrary gauge-factor; because by the removal of the rigid connection between tetrads at

different points the gauge-factor necessarily becomes an arbitrary function of position. But then the infinitesimal linear transformation dE of the ψ, which corresponds to the infinitesimal rotation $d\Omega$, is not fully determined, but can be extended by an arbitrary pure imaginary multiple $i.df$ of the unit matrix. For the unique determination of the covariant differential $\delta\psi$ of ψ one needs, in addition to the metric in the neighbourhood of the point P, such a df for every line element $PP' = (dx)$ radiating from P. In order that $\delta\psi$ still depend linearly on dx

$$df = f_p(dx)^p$$

must be a linear form in the components of the line-element. If one replaces ψ by $e^{i\lambda}\psi$ one must, according to the covariant differential formalism, replace df by $df - d\lambda$ at the same time.

The result of this is that the term

$$\frac{1}{\epsilon}f(\alpha)s(\alpha) = \frac{1}{\epsilon}f(\alpha).\psi^* S(\alpha)\psi = f_p.\psi^*\mathfrak{S}^p\psi \tag{26}$$

must be added to the action-density $\mathfrak{m}$. From now on $\mathfrak{m}$ will denote the complete action density. We then have gauge-invariance in the sense that the action remains unchanged under the replacements

$$\psi \quad \text{by} \quad e^{i\lambda}\psi, \qquad f_p \quad \text{by} \quad f_p - \frac{\partial\lambda}{\partial x_p},$$

where λ is understood to be an arbitrary function of position. It is exactly in the manner described by (26) that the electromagnetic potential interacts with matter according to experiment. This justifies the identification of the quantities f_p introduced here with the electromagnetic potentials. The proof is complete if we show that, conversely, the f_p-field is influenced by the matter according to the laws that are shown by observation to be valid for the electromagnetic field.

$$f_{pq} = \frac{\partial f_q}{\partial x_p} - \frac{\partial f_p}{\partial x_q}$$

is a gauge-invariant anti-symmetric tensor and

$$\mathfrak{l} = \frac{1}{4}f_{pq}\mathfrak{f}^{pq} \tag{27}$$

the scalar density characteristic of the Maxwell field. The Ansatz

$$\mathfrak{h} = \mathfrak{m} + a\mathfrak{l} \tag{28}$$

(a a numerical constant) yields by variation of f_p the Maxwell equations with

$$-\mathfrak{s}^p = -\psi^* \mathfrak{S}^p \psi \tag{29}$$

as the electrical 4-current density.

The gauge-invariance is closely connected with the conservation of electric charge. Since $\mathfrak{h}$ is gauge-invariant $\delta \int \mathfrak{h} dx$ must vanish identically when for fixed $e^p(\alpha)$ the ψ and f_p are varied according to

$$\delta\psi = i\lambda\psi, \qquad \delta f_p = -\frac{\partial\lambda}{\partial x_p},$$

λ being an arbitrary function of position. This produces a relation between the material and electromagnetic equations which is identically satisfied. If we know that the field-equations for the matter (in the narrow sense) are valid it follows that

$$\delta \int \mathfrak{h}.dx = 0,$$

when the f_p alone are varied according to $\delta f_p = -\partial\lambda/\partial x_p$. On the other hand the same result follows from the equations of electromagnetism for the infinitesimal variation $\delta\psi = i\lambda.\psi$ of the ψ alone. If $\mathfrak{h} = \mathfrak{m} + a\mathfrak{l}$ one obtains in each case

$$\int \delta\mathfrak{h}.dx = \pm \int \psi^* \mathfrak{S}^p \psi \frac{\partial\lambda}{\partial x_p}.dx = \mp \int \lambda \frac{\partial \mathfrak{s}^p}{\partial x_p} dx.$$

We found earlier an analogous situation for the conservation of energy-momentum and angular momentum. They connected the field-equations for matter in the wider sense with the gravitational equations and corresponded to invariance with respect to coordinate transformations or equivalently with respect to independent rotations of the local tetrads at different points. From

$$\frac{\partial \mathfrak{s}^p}{\partial x_p} = 0 \tag{30}$$

it follows that the flow of the vector-density s^p through a three-dimensional section of space-time, in particular through a section (22)

$$\mathbf{l} = \int \mathfrak{s}^o d\xi \tag{31}$$

is independent of the position of the section, respectively t. Not only the integral, but each separate integral-element has an invariant meaning; however the sign depends on which direction is taken as the positive direction of the three-dimensional section. For $\mathfrak{s}^0 d\xi$ to be a space-like probability-density the hermitian form

$$e^o(\alpha)\psi^* S(\alpha)\psi \tag{32}$$

of ψ_1, ψ_2 must be definite. One easily finds that this is the case if (22) is actually a space-like section through P i.e. if the line-elements originating in P are space-like. In order to obtain the positive sign in (32) the sections $x_0 =$ constant, ordered according to increasing x_0, should follow the future direction determined by the vector $\mathbf{e}(0)/i$. With this natural restriction of the coordinate-system the sign of the current is determined and the invariant (31) is normalized in the usual way by the condition

$$\mathfrak{I} \equiv \int \mathfrak{s}^o d\xi = 1 \tag{33}$$

The constant a that connects $\mathfrak{m}$ with $\mathcal{L}$ is then a pure number $= ch/e^2$ (reciprocal of the fine-structure constant).

We treat ψ_1, ψ_2; f_p; $e^p(\alpha)$ as the mutually independent quantities to be varied. On account of the additional term in (26) the energy density $\mathfrak{t}_q^p$ coming from $\mathfrak{m}$ must be extended by

$$f_p \mathfrak{s}^q - \delta_p^q (f_r \mathfrak{s}^r).$$

In special relativity this leads to the energy being associated with the operator

$$H = \sum_{p=1}^{3} S^p \left(\frac{1}{i} \frac{\partial}{\partial x_p} + f_p \right)$$

since it has the value

$$\int \psi^* H \psi d\xi.$$

Of course, the matter equations then take the form

$$\left(\frac{1}{i} \frac{\partial}{\partial x_o} + f_o \right) \psi + H\psi = 0 \quad \text{and not} \quad \frac{1}{i} \frac{\partial \psi}{\partial x_o} + H\psi = 0,$$

as was formerly assumed in quantum-mechanics. Naturally, the matter-energy must be supplemented by the electromagnetic energy, for which Maxwell's classical expression is still valid.

As far as physical dimensions are concerned, it is natural in general relativity to take the x_p to be pure numbers. The quantities that appear are not only invariant with respect to scale-changes but with respect to arbitrary variations of the x_p. If we make the change $\mathbf{e}(\alpha) \to b\mathbf{e}(\alpha)$ by multiplication with a constant b then to preserve the normalization the ψ must be changed to $b^{3/2}\psi$. Thereby $\mathfrak{m}$ and $\mathfrak{l}$ are not changed and are therefore pure numbers. In contrast, $\mathfrak{g}$ acquires the factor $1/b^2$ so that κ is the square of a length d. κ is not identical with Einstein's gravitational constant but is obtained from it by multiplication with $2h/c$. The constant d is much smaller than the atomic order of magnitude, it is $\sim 10^{-32}$. Thus here also gravitation will be significant only for astronomical problems.

Thus, if we ignore gravity, the field-equations contain no dimensional atomic constant. In the two-component theory there is no place for a term in the action like the mass-term in the Dirac theory [6]. But one knows how the mass can be introduced on the basis of the conservation laws. One assumes that in the "empty environment" of the particle, outside a certain space-time tube whose $x_0 =$ constant section is finite, the $\mathfrak{t}_p^q$ vanish and the $e^p(\alpha)$ take the constant values of special relativity. Then

$$J_p = \int \left(\mathfrak{t}_p^o + \frac{1}{\kappa}\mathfrak{v}_p^o\right) d\xi$$

are the components of a four-vector which is constant in time and is not influenced by the choice of coordinate system or the local tetrad. The normal coordinate system can be fixed more precisely by the condition that the three-momentum (J_1, J_2, J_3) vanish; then $-J_0$ is the invariant and simultaneously the constant mass of the particle. It is then required that the value of this mass be given once and for all.

In addition to the above-mentioned theory of the electromagnetic field, which I regard as correct because it originates so naturally in the arbitrariness of the gauge-factor in ψ and makes the connection between the experimentally observed gauge-invariance and charge-conservation understandable, there is another possible theory that connects electricity with gravitation. The term (26) has the same form as the second part of in (17); $\phi(\alpha)$ plays in the latter the same role as $f(\alpha)$ in the former. Hence one might expect that matter and gravitation, ψ and $e^p(\alpha)$, are already sufficient to explain electromagnetic phenomena, by considering the $\phi(\alpha)$ as the electromagnetic potentials. The dependence of these quantities on the $e^p(\alpha)$ and their first derivatives is such that they are invariant with respect to arbitrary coordinate transformations. But with respect to rotations of the tetrads the $\phi(\alpha)$ transform like fixed vectors of the tetrad only if the tetrads are subjected to the same rotations at all points. If one ignores the matter and considers only the relationship between electromagnetism and gravitation, one arrives at a theory of electromagnetism which is exactly of the same kind as that recently attempted by Einstein. However, in the present case the distant-parallelism is only simulated.

I have persuaded myself that the latter Ansatz, which at first sight is perhaps attractive, cannot yield the Maxwell equations. Furthermore the gauge-invariance would remain quite puzzling; the electromagnetic potentials and not merely the field-strengths would have a physical meaning. Thus I believe that this idea is a red-herring and that we should place much more trust in the hint that is provided by gauge-invariance: that electromagnetism is an accompanying phenomenon of the material wave-field and not of gravitation.

Palmer Physical Laboratory, Princeton University, 19, April 1929

[1] E. Wigner, ZS. f. Phys. **53**, 592, 1229; among others.

[2] In formal agreement with Einstein's recent work on Gravitation and Electricity, Sitzungsber. Preuss. Ak. Wissensch. 1928, p. 217, 224; 1929, p.2. Einstein uses the letter h instead of e.

[3] See for example H. Weyl, Space-Time-Matter 5th Ed., p. 233ff. (cited as RZM). Berlin 1923

[4] See RZM, p. 119f.

[5] See RZM, p. 272f.

[6] Proc. Roy. Soc. (A) **117**, 610.