Additional copies may be obtained from
Gravity Research Foundation
New Boston, N. H.

Gravity and the Nature of Fundamental Particles

by
Lloyd Motz
Rutherfurd Observatory
Columbia University

1960

Summary

Up until the present time it has not been possible to establish a relationship between the macroscopic gravitational theory and the properties of fundamental particles. Ultimately, however, if one is to attain a thorough understanding of gravitation it must be through a correct theory of the structures of the fundamental particles. For that reason the present essay is devoted to an analysis of this point.

Since the macroscopic gravitational theory is based upon non-Euclidean geometries, one may hope to acquire an understanding of fundamental particles by a similar approach. In this essay the *Weyl theory of gauge invariance* is applied to obtaining a geometrical theory of electrons and protons, and it is shown that their structure can be understood in terms of localized non-Euclidean geometries.

Introduction

The purpose of this essay is to present a new theory of fundamental particles which is a natural extension of the Einstein gravitational theory and which at the same time achieves a unification of electromagnetism and gravitation along the lines first proposed by Hermann Weyl and later extended by A. S. Eddington[1]. By making use of the concept of gauge invariance as introduced by Weyl, we shall see that it is possible to picture fundamental particles such as electrons and protons in terms of localized non-Euclidean manifolds in which the radius of curvature is of the order of the Compton wave-length for the particular particle under consideration.

The great developments such as the special and the general theories of relativity that have occurred in twentieth century physics have been the result of the extension of the concept of invariance to include manifolds beyond those covered by the transformations in classical physics. Thus the Einstein-Lorentz transformations extended the notion of the invariance of Galilean systems to four-dimensional manifolds in which time as a fourth dimension was introduced on an equal footing with the three dimensions of space. The general theory went still further by postulating the invariance of the space-time line element to the most general type of transformation of coordinates and not just to the Lorentz group, thus achieving a synthesis of non-Euclidean geometry and gravitation in the large.

Considering the successes of these principles of invariance, one would be naturally inclined to obtain a still greater unification of the laws of nature by extending the concept of invariance still further, and this is what Weyl attempted to do by introducing the principle of gauge invariance. He hoped by doing this to obtain a unified field theory in which the electromagnetic field would be generated by the imposition of gauge invariance just as the gravitational field is generated by the condition of coordinate invariance. Although Weyl succeeded in relating his principle of gauge invariance to a second order anti-symmetric tensor which he identified with the electromagnetic field, there were some serious objections to his theory which made it unacceptable. In this essay we shall show that the principal objection to the Weyl theory can be removed if one imposes a uniqueness condition on the gauge which leads to the Bohr-Sommerfeld quantum conditions. An additional constraint which we shall impose on the integrability of the length of an interval will lead to the Lorentz pondermotive force.

We shall show first that we can derive the Lorentz expression for the pondermotive force on a charge in an electromagnetic field if we impose upon the Weyl theory the condition that the charge must move in such a way that the change in its dimensions along a given path (resulting from the change of gauge along this path) must be a minimum. Secondly, we shall show that the Bohr-Sommerfeld quantum integral follows directly if we impose the condition that the dimensions of a charged particle, except for a possible change of phase, must return to their initial values when the particle moves around a closed orbit. We shall see that this condition will eliminate one of the most serious objections to the Weyl theory arising from the non-integrability of length that is a consequence of this theory. Finally, we shall demonstrate that the generalized Lagrangian density obtained from (1) leads directly to the second order Dirac equation for an electron in an electromagnetic field provided that we assume that the Gaussian curvature in the neighborhood of a particle is given by the reciprocal of its Compton wave-length.

The Lorentz Pondermotive Force

The Weyl theory of gauge invariance arose out of the concept that lengths at different places cannot be compared because of the change of gauge that takes place as one moves from point to point in a space-time continuum. Since the gauge was assumed to be determined by a vector field κ_μ, comparison of lengths at different places would be ambiguous because the result of the comparison would depend on the path taken in going from one point to the other. Although this theory introduced a four-vector κ_μ (to be identified with the electromagnetic vector potential) into a description of the world quite naturally, the non-integrability of length which it brought with it led to apparently insurmountable difficulties concerning the structure of atoms.

Thus the objection was raised that according to the Weyl theory the natural frequency of an atom at a point in space-time should depend on the path the atom took to reach that point. This objection was met by introducing the assumption that although lengths and frequencies depend on the path taken, the effect is much too small to be measurable in actual physical phenomina. This, however, is not a satisfactory way out of the difficulty since the ambiguity is still present in the theory. It is possible to eliminate this ambiguity without destroying the content of the Weyl theory by imposing the condition that the measurable physical dimensions of a particle shall be integrable along the path of its motion. We must note that this is not the same thing as imposing the condition that the gauge be integrable along a path. This latter condition is much too restrictive and would result in the vanishing of the curl of the four-vector κ_μ and hence to the vanishing of the electromagnetic field so that our theory would be empty.

The possibility of imposing integrability on the physically meaningful dimensions and yet not on the gauge arises from the fact that the dimensions must be treated as complex quantities so that they have arbitrary phase factors associated with them. Since these phases need not be integrable, the gauge will not be integrable either, with the result that the content of the theory will remain while the ambiguity is eliminated. We shall come back to this point in our discussion of the Bohr-Sommerfeld quantum condition, but now we shall consider what constraints may be imposed on the motion of a particle without modifying the non-integrability of the gauge at all.

To see what we must do we shall start from Weyl's fundamental assumption that if a length A is displaced from a point x_μ to a nearby point $x_\mu + dx_\mu$, then it suffers a change in length determined by the equation

$$d \log A = \kappa_\mu d x_\mu \qquad (1)$$

where κ_μ is a vector-field. Let us suppose now that we have a particle which moves from some point, P_1, in our space-time continuum to some other point, P_2, along a physically permissible path. What constraint can we impose upon this motion that will be physically significant and yet which will not violate the basic assumptions of the Weyl theory? It is reasonable to assume that a particle will tend to retain its dimensions in so far as possible as it moves along its path. We shall therefore impose the condition that the particle will move along that particular path connecting the end points P_1 and P_2 which results in the smallest change in its dimensions. In other words we shall assume that

$$\int_{P_1}^{P_2} d \log A = \int_{P_1}^{P_2} \frac{dA}{A} = \int_{P_1}^{P_2} \mu_\mu \, dx_\mu = \text{a minimum.} \qquad (2)$$

We have, then, as an additional constraint on the motion of the particle the condition that

$$\delta \int_{P_1}^{P_2} \kappa_\mu \, dx_\mu = 0 \qquad (3)$$

for all variations of the permissible path in which the end points are kept fixed.

If ds is the element of path length along a geodesic between two fixed points, it must satisfy the stationary condition obtained from the general theory of relativity

$$\int_{P_1}^{P_2} \delta \, (ds) = 0 \qquad (4)$$

However, the variation in the integrand of (4) may no longer be taken as arbitrary since only those variations are permissible which are governed by equation (3). We may treat this situation in the usual way by varying (3) and incorporating this variation into (4) by means of Lagrangian multipliers.

The mathematical operations can be carried out quite easily, and if we introduce the second order anti-symmetric tensor $\mathfrak{F}_{\mu\nu}$ defined by

$$\mathfrak{F}\mu\nu = \frac{\partial \kappa_\mu}{\partial X_\nu} - \frac{\partial \kappa \nu}{\partial X_\mu} \qquad (5)$$

and the four-velocity V_μ, we obtain the equation

$$\frac{d^2 X_\alpha}{dS^2} + \Gamma^\alpha_{\mu\sigma} \frac{dX_\mu}{dS} \frac{dX_\sigma}{dS} = L \mathfrak{F}^\alpha_\mu V_\mu. \qquad (6)$$

where the $\Gamma^\alpha_{\mu\sigma}$ are the affine connections of the geometry defined in the usual manner and L is the Lagrangian multiplier.

Since κ_μ will later be related to the vector potential A_μ by the equation $\kappa_\mu = (i/\hbar)(e/c) A_\mu$, the Lagrangian multiplier, L, must be chosen equal to $-i(\hbar/mc)$ in order to make equation (6) dimensionally correct. The letters e, $\hbar$, m and c have their usual meanings, viz: electron charge, Planck constant, electron mass, and the speed of light, respectively.

Since the left hand side of (5) represents the acceleration of a particle, the right hand side must be the force acting on it and this is, indeed, the Lorentz force if L is chosen as above. We see that this is a natural extension of the equation of a geodesic which describes the motion of a free particle in a gravitational field.

The Bohr-Sommerfeld Quantum Integral

We have already noted that the non-integrability of length which follows from the Weyl theory brings certain objectionable features with it which cast doubt on the entire theory. We must therefore try, in so far as is possible, to eliminate these features but not at the expense of the physical content of the theory. We may do this if we note that the quantity $\log A$ is, in general, not real so that we may write $A = \mathfrak{A} e^{i\phi}$, where ϕ is a real number. The arbitrary phase factor will have no effect on the physically meaningful lengths in nature since these are to be obtained from the mathematical quantities A by taking absolute values. We now have from equation (1) the result

$$d \log \mathfrak{A} + i d\phi = \kappa_\mu \, dx$$

or

$$d \log \mathfrak{A} = \kappa_\mu \, dx\mu - i d\phi \qquad (7)$$

We shall now impose the condition that $\log \mathfrak{A}$ shall be integrable along any permissible path of a particle but that $\log A$ need not be. If we now consider a particle moving in a closed orbit in the field we see that we must have

$$\oint d \log \mathfrak{A} = \oint \kappa_\mu \, dx_\mu - i \oint d\phi = 0$$

or

$$\oint \kappa_\mu \, dx_\mu = i \oint d\phi. \qquad (8)$$

We have complete freedom in terms of our theory as to the change that ϕ must suffer when our particle moves once around its orbit, but it is most natural to assume the change will be such as to have as small an effect as possible on A, and this will obviously be the case if ϕ changes exactly by an integral multiple of 2π. We therefore have from (8) the additional constraint on the motion of the particle given by

$$\oint \kappa_\mu \, dx_\mu = 2\pi i n \qquad (9)$$

where n is any integer.

If we now replace κ_μ by its definition in terms of the vector potential A_μ as given in the final paragraphs of the previous section, we have

$$(e/c) \oint A_\mu \, dx_\mu = nh \qquad (10)$$

and this can be shown to be just the Bohr-Sommerfeld condition for the motion of a charged particle in an electromagnetic field defined by the four-vector A_μ.

The Structure of Matter

To use the theory of gauge invariance to obtain some insight into the structure of matter we must construct a Lagrangian that is relativistically invariant and also invariant to transformation of gauge. Since this Lagrangian must contain the properties of the electromagnetic field as well as of the gravitational field, it must be constructed of a second order tensor containing an anti-symmetrical as well as a symmetrical part. The general theory of relativity gives us the Einstein-Ricci symmetrical curvature tensor, $G_{\mu\nu}$, and from it we can obtain a gauge invariant tensor by imposing the Weyl condition. We are thus led to the tensor

$$^*G_{\mu\nu} = G_{\mu\nu} - (\kappa^{\alpha}_{;\alpha} - 2\kappa_\alpha \kappa^\alpha) g_{\mu\nu} - 2\kappa_\mu \kappa_\nu + (\kappa_{\mu,\nu} + \kappa_{\nu,\mu}) - 2\mathfrak{F}_{\mu\nu} \qquad (11)$$

where the $g_{\mu\nu}$ are the components of the metric tensor, and $\kappa_{\mu,\nu} = \dfrac{\partial \kappa_\mu}{\partial X_\nu}$, and $\mathfrak{F}_{\mu\nu}$ is defined by (5).

This tensor, which is the sum of a symmetrical and antisymmetrical part, was used by the present author[2] to construct a generalized Lagrangian from which a set of Maxwell-Lorentz equations for an electron was derived. To do this it was necessary to linearize the Lagrangian by introducing Dirac matrices so that the electromagnetic field automatically generated charged particles with spin.

To obtain a model for a fundamental charged particle such as an electron it is necessary to start with an invariant linear Lagrangian from which the Dirac equation can be obtained. To do this we must multiply (11) by another second order contravariant tensor having the same symmetry properties as $^*G_{\mu\nu}$, and the simplest such tensor is the sum of the metric and the angular momentum tensors, that is, the general spin tensor.

$$M^{\mu\nu} = \tfrac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu) + \tfrac{1}{2} i (\gamma^\mu \gamma^\nu \gamma^\nu \gamma^\mu) = g^{\mu\nu} + \sigma^{\mu\nu}) \qquad (12)$$

where the γ's are the usual Dirac spin matrices. Since the first term on the right hand side is symmetric and the second one is anti-symmetric, we obtain

$$M^{\mu\nu} \, ^*G_{\mu\nu} = 4\lambda + \sigma^{\mu\nu} \mathfrak{F}_{\mu\nu}, \qquad (13)$$

where

$$\lambda = (\tfrac{1}{4}) Q \, [G - 6(\kappa^{\alpha}_{\alpha} - \kappa^\alpha \kappa_\alpha)]. \qquad (14)$$

Q is a scale factor and G is the Gaussian curvature at the position of the particle.

We shall now take as our Lagrangian density the expression

$$\mathfrak{L} = \overline{\Psi}(X)[4\lambda + \sigma^{\mu\nu}\mathfrak{F}_{\mu\nu}] \, \Psi(X), \qquad (15)$$

where $\Psi(X)$ is the four-component Dirac spinor that has to be varied and $\overline{\Psi}(X)$ is its adjoint. This leads to the following variational principle:

$$\delta \int \overline{\Psi}(X)[4\lambda + \sigma^{\mu\nu}\mathfrak{F}_{\mu\nu}]\Psi \quad (X) \, d\tau = 0 \qquad (16)$$

On carrying out the variation with respect to the spinors, we obtain the equation

$$[\sigma^{\mu\nu}\mathfrak{F}_{\mu\nu} - \frac{6}{Q}(\kappa^{\alpha}_{;\alpha} - \kappa^\alpha \kappa_\alpha)]\Psi(X) = -\frac{G}{Q}\Psi(X) \qquad (17)$$

where we have introduced the expression (14) for λ.

Since we have not limited ourselves to any particular choice of gauge in (17) we may transform to any gauge we may desire by applying the combined transformations

$$\kappa_\nu \rightarrow \kappa_\nu - \frac{\partial S}{\partial X}$$

$$\Psi(X) \rightarrow e^{-S}\Psi(X) \qquad (18)$$

where S is a pure imaginary, dimensionless quantity since the 4-vector κ_ν is a pure imaginary with the dimensions of a reciprocal length.

Using some elementary transformation we obtain:

$$\left[\sigma^{\mu\nu}\mathfrak{F}_{\mu\nu} - \frac{6}{Q}\left(\frac{\partial}{\partial X_\nu} - \kappa_\nu\right)\left(\kappa^\nu - \frac{\partial}{2X_\nu}\right)\right]\Psi(X) = -\frac{G}{Q}\Psi(X). \qquad (19)$$

We now replace the 4-vector κ_ν by the electromagnetic 4-vector A_ν. Since these two vectors differ in their dimensions by a charge, we shall write $\kappa_\nu = (ie/\hbar c)A_\nu$, so that we get $Q = -2\,ie/\hbar c$. If we introduce these expressions into (19) and incorporate the numerical factor 2/3 into the field strengths $\mathfrak{F}_{\mu\nu}$, we obtain the equation

$$\left[\frac{ie}{\hbar c}\sigma^{\mu\nu}\mathfrak{F}_{\mu\nu} + \left(\frac{1}{i}\frac{\partial}{\partial X_\nu} - \frac{e}{\hbar c}A_\nu\right)\left(\frac{1}{i}\frac{\partial}{\partial X_\nu} - \frac{e}{\hbar c}A^\nu\right)\right]\Psi(X) = \frac{G}{6}\Psi(X) \qquad (20)$$

We note that the right hand side of (20) is essentially the square of the curvature of space at a point occupied by a particle, in other words the reciprocal of the square of some fundamental length associated with the particle. Since there are only two fundamental lengths that can be constructed from the constants e, $\hbar$, m, and c, and one of them, the classical radius of the electron, is non-quantum mechanical in nature, we shall identify the right hand side of (20) with the other one. In other words, we shall assume that $G/6$ is equal to $(\hbar/mc)^{-2}$. If we introduce this into (20) and multiply the entire equation by $\hbar^2$, we are led to the equation

$$\left[i\frac{e\hbar}{c}\sigma^{\mu\nu}\mathfrak{F}_{\mu\nu} + \left(\frac{\hbar}{i}\frac{\partial}{\partial X_\nu} - \frac{e}{c}A_\nu\right)\left(\frac{\hbar}{i}\frac{\partial}{\partial X^\nu} - \frac{e}{c}A^\nu\right)\right]\Psi(X) = m^2c^2\Psi(X). \qquad (21)$$

Discussion

In deriving the second order Dirac equation (21) for an electron in an electromagnetic field we have had to make an important assumption concerning the geometry of space-time in the neighborhood of an electron, viz., that the Gaussian radius of curvature of space in the neighborhood of a particle is given by the Compton wave-length.

There appears, however, to be justification for assigning so important a geometrical role to the Compton wave-length because of two properties, one relativistic and the other quantum mechanical in nature, associated with it. There is a theorem in special relativity which states that a system with positive energy and inner angular momentum $\hbar$ and with a given rest mass m must have a finite extension which cannot be smaller than $\hbar/mc$.[3] In the sense that this represents the maximum accuracy with which the position of an electron can be determined, this theorem appears to be closely related to a similar result derived from the Compton

effect.

Because the determination of the position of an electron must be carried out by observing photons after they have been scattered by the electron, it follows that the Compton effect will introduce an uncertainty in the position so determined because the wave-length of the scattered radiation is different from that of the incident radiation. In fact, even if the incident wave-length is zero, the wave-length of the scattered photon will be of the order of the Compton wave-length. Pauli[4] has shown that the minimum value of the scattered wave-length will be of the order of $(\hbar/mc)[1 - (v/c)^2]^{1/2}$ if v is the speed of the electron. For an electron at rest this is just the Compton wave-length.

One further point may be mentioned in connection with the assumption we have made concerning the Compton wave-length. Such an assumption implies that the geometry of space-time in the neighborhood of an electron (neglecting electromagnetic effects of external fields) must be governed by an equation of the form[5]

$$G^{\mu\nu} = 3(mc/\hbar)^2 g^{\mu\nu} \qquad (22)$$

If we neglect gravitational effects, this law leads to a line element of the form

$$ds^2 = -(1-\alpha r^2)^{-1}dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\phi^2 + (1-\alpha r^2)\,dt^2, \qquad (23)$$

where α is the reciprocal of the square of the Compton wave-length.

We see from this that the coefficient of dr^2 vanishes for r equal to the Compton wave-length, suggesting that this radius represents a kind of impassable barrier. We may also see in this a possible mechanism for the capture of a photon by an electron if we note that the path of the photon is always given by setting (23) equal to zero. However, at a distance from the electron equal to the Compton wave-length ds becomes infinitely large unless $dr = 0$. In other words, a photon can move in the neighborhood of an electron only if it has no radial component in its motion. Such photons, having no radial velocity relative to the electron, must circulate around the electron, and may therefore be considered as being captured.

References

(1) A. S. Eddington, *The Mathematical Theory of Relativity*, page 204. Cambridge University Press, Cambridge 1923

(2) L. Motz, *Physical Review*, 89:60, 1953

(3) C. Moller, *Relativity*, page 177. Oxford Press 1952

(4) W. Pauli, *Wellenmechanik*, page 91. J. W. Edwards 1946

(5) A. S. Eddington, *op. cit.*, page 100