Gauge Invariance and Classical Electrodynamics

LLOYD MOTZ
Rutherfurd Observatory, Columbia University, New York, New York
(Received February 25, 1952; revised manuscript received September 30, 1952)

A Lagrangian for deriving the Maxwell-Lorentz field equations is obtained by starting from a general tensor of the second rank $\mathcal{G}_{\mu\nu}$, which is expressed as the sum of a symmetric and an antisymmetric part $\mathcal{G}_{\mu\nu}=\lambda g_{\mu\nu}+F_{\mu\nu}$. The symmetric part is identified with the metric tensor and the antisymmetric part with the electromagnetic field. A relationship between this second rank tensor and the gauge invariant Ricci-Einstein tensor is established by means of the gauge invariant theories of Weyl and Eddington. This relationship leads directly to the Klein-Gordon relativistic wave equation for a point charge moving in an electromagnetic field provided the function λ is properly chosen.

The Lagrangian density is defined as the quantity $(-|\mathcal{G}_{\mu\nu}|)^{\frac{1}{2}}$, where $|\mathcal{G}_{\mu\nu}|$ is the determinant associated with the tensor $\mathcal{G}_{\mu\nu}$. This choice is made for the Lagrangian density since $(-|\mathcal{G}_{\mu\nu}|)^{\frac{1}{2}}d\tau$ is the simplest generalized invariant volume element. Since the Lagrangian density is nonlinear and irrational as it stands, it is first rationalized by means of the Dirac matrices. If the Lagrangian that is obtained in this way is varied with respect to the vector and scalar potentials, one obtains a set of Maxwell-Lorentz equations for a charge-current distribution that is defined in terms of the field potentials. These equations are almost identical with those recently obtained by Dirac in his new classical electrodynamics. It is shown from the field equations that the velocity of the charge distribution is given by the relation

$$\mathbf{v}=(\mathbf{A}/\phi)c,$$

where $\mathbf{A}$ is the vector potential, ϕ is the scalar potential, and c is the speed of light. This is identical with the result obtained from the retarded potentials of a point charge. For the static case it is shown that the field equations lead to solutions for the fields and potentials that are finite everywhere, and the self-energy of the point charge is finite. The classical radius of the electron, the Compton wavelength and the fine structure constant come into the theory quite naturally.

INTRODUCTION

It is generally believed that the infinity difficulties associated with the self-energy of the electron and the electromagnetic field are due to the fact that thus far it has been impossible to introduce into the theory a fundamental length of the order e^2/mc^2 in a satisfactory and relativistically invariant manner. Although recent developments in the quantum electrodynamics have produced a prescription for subtracting these infinities in an unambiguous way so that one is left with a result that is finite, the fundamental problem of accounting for the finite mass, charge, and self-energy of the electron still remains.

Since the introduction of a point charge (as demanded by the condition of relativistic invariance) already brings with it infinities in the classical electromagnetic theory, one should attempt to remove the infinity difficulties at this stage before passing over to a quantized theory. In the past a number of attempts have been made to achieve this result but without much success. Most of these older theories, of which that of Born and Infeld[1] is a good example, have led to nonlinear field equations which are extremely cumbersome and difficult to work with.

Recently Dirac[2] has introduced a new form of classical electrodynamics by adding to the old Lagrangian a term which is quadratic in the field potentials and has the effect of destroying the gauge invariance of the theory. Although Dirac is able to obtain the field equations for a charge-current distribution from this Lagrangian, the introduction of the additional term is performed rather artificially.

In the present paper, a new approach to this fundamental problem is suggested which enables one to obtain a new set of field equations which are linear and yet which lead to solutions for the field intensities which are everywhere finite. The charge-current distribution appears in a natural way, and for weak fields the Lagrangian and the Maxwell-Lorentz field equations reduce to those obtained by Dirac.

The essential departure from the Born-Infeld procedure that is taken in this paper is to replace the irrational Lagrangian they adopted, namely, the square root of a quadratic form, by a rational one that can be obtained from a tensor of the second rank by introducing the Dirac matrices. This second-rank tensor is not chosen arbitrarily but is determined from the condition of gauge invariance as introduced by Weyl and others. With this choice the charge-current distribution is automatically introduced and related in a simple way to the Gaussian curvature of the space or the mechanical density of the distribution. If the symmetric part of the second-rank tensor is properly chosen, the usual condition of gauge invariance of the electromagnetic potentials leads to the Klein-Gordon relativistic wave equation for the charge distribution.

THE GENERALIZED VOLUME AND THE LAGRANGIAN DENSITY

Since any tensor of the second rank can be written as the sum of a symmetric and an antisymmetric part, we shall take as the starting point of our theory the asymmetric second rank tensor

$$\mathcal{G}_{\mu\nu}=\lambda g_{\mu\nu}+F_{\mu\nu}, \tag{1}$$

where λ is some invariant function of the space-time coordinates which we shall specify later, $g_{\mu\nu}$ is to be identified with the metric tensor, and $F_{\mu\nu}$ is an anti-

[1] M. Born and L. Infeld, Proc. Roy. Soc. (London) **A144**, 425 (1934).
[2] P. A. M. Dirac, Proc. Roy. Soc. (London) **A209**, 291 (1951).

symmetric tensor that defines the electromagnetic field. The choice of (1) is not entirely arbitrary but is suggested by certain features of the gauge invariant theory developed by Weyl. We shall see, in particular, that the function λ can be chosen in such a way that one is led directly to the Klein-Gordon wave equation.

It is interesting to note that the property of gauge invariance is a feature not only of the electromagnetic potentials but also of the wave functions associated with particle fields. One might therefore expect that any theory which creates an electromagnetic potential by means of a gauge transformation should at the same time create a particle field. We shall see in what follows that this is true of the Weyl theory of gauge invariance. The possibility of deriving particle wave functions from the Weyl theory of gauge invariance was first investigated by London,[3] who identifies the integral of the four-vector obtained from Weyl's theory with the phase of the Schrödinger wave function.

The Weyl theory of gauge invariance is based on the notion that it is not possible to compare lengths at different places in space-time because the result of the comparison will depend on the path taken in bringing the two lengths together. In other words, if we consider the square of the line element $ds^2=g_{\mu\nu}dx_\mu dx_\nu$, only the relative values of $g_{\mu\nu}$ have physical significance. This theory has been criticized because it would seem to indicate that the natural frequency of an atom at a point in space-time should depend on the path taken by the atom to reach the point. This criticism can be answered in various ways.

It may, for example, be argued that although the frequency of an atom will depend on the path taken by the atom, the effect is so small as to be entirely beyond experimental verification for the conditions ordinarily met with in the laboratory. In other words, the effect may be of great theoretical significance and yet of no importance experimentally. The criticism may also be answered by stating that the comparison of lengths does not depend on the parallel displacement of the lengths from point to point but is a property of the geometry at the particular point where the lengths happen to be. This means that at each point of space-time a particular gauge exists which is determined only by the magnitude of the field intensities at that point. Whenever a particle reaches some point, it automatically adjusts its dimensions to the appropriate gauge regardless of the path it traversed to reach the point. As a final argument we note that ds is a complex quantity, in general, so that we may say that the variation in ds as one moves from point to point occurs in its phase and not in its absolute magnitude. Thus the physical quantities associated with the line element are independent of the path taken, and only the physically meaningless phase is altered.

If a small space-time interval at some arbitrary point (x_μ) is transferred by parallel displacement to a neighboring point $(x_\mu+dx_\mu)$, its length will change from a value l to a value $l+dl$ because of the change in gauge in going from the first point to the second. This change in length can be expressed by the formula

$$d(\log l)=\kappa_\mu dx_\mu, \tag{2}$$

where the κ_μ are the components of a four-vector. We note that these quantities have the dimensions of a reciprocal length. In the original form of the Weyl theory, these quantities were identified with the electromagnetic potentials, but we shall not do that at this stage.

Let us now consider a dimensionless invariant space-time function S which is so defined that $(\exp S)$ describes a new gauge system. This new gauge system is introduced by altering the length of our unit at each point in the ratio $(\exp S)$. We thus introduce into the theory a scalar field S in addition to the vector field κ_μ.

Let us now suppose that the gauge of our system is altered in the manner described above and consider what changes are introduced as a result of this. If the starred quantities represent the new values after the change of gauge, we obtain

$$\begin{aligned} &\text{line element:} && ds^*=ds(\exp S);\\ &\text{metric tensor:} && g^*_{\mu\nu}=g_{\mu\nu}(\exp 2S). \end{aligned} \tag{3}$$

We also have, from (2),

$$\begin{aligned} \kappa_\mu{}^* dx_\mu=d(\log l^*)&=d\log(l\exp S)\\ &=d(\log l)+(\partial S/\partial x_\mu)dx_\mu, \end{aligned} \tag{4}$$

or

$$\kappa_\mu{}^*=\kappa_\mu+(\partial S/\partial x_\mu).$$

If we now define the antisymmetric tensor $\mathfrak{F}_{\mu\nu}$ as the curl of the vector κ_μ,

$$\mathfrak{F}_{\mu\nu}=(\partial\kappa_\mu/\partial x_\nu)-(\partial\kappa_\nu/\partial x_\mu), \tag{5}$$

we see that $\mathfrak{F}_{\mu\nu}{}^*=\mathfrak{F}_{\mu\nu}$ so that $\mathfrak{F}_{\mu\nu}$ is independent of the gauge.

The idea behind the Weyl theory was to introduce into the formalism of general relativity only those tensors that are gauge invariant. Eddington calls these quantities in-tensors. Since the affine connections are functions of the metric tensor and its derivatives, it is clear that the physically important Riemann-Christoffel tensor is not an in-tensor. It is, however, fairly easy to generalize this tensor so that it is gauge invariant. If one does this and contracts it,[4] one obtains the second-rank in-tensor

$$\begin{aligned} {}^*G_{\mu\nu}=G_{\mu\nu}-(\kappa^\alpha{}_{,\alpha}-2\kappa_\alpha\kappa^\alpha)g_{\mu\nu}-2\kappa_\mu\kappa_\nu\\ -(\kappa_{\mu,\nu}-\kappa_{\nu,\mu})-2\mathfrak{F}_{\mu\nu}, \end{aligned} \tag{6}$$

where $G_{\mu\nu}$ is the usual Einstein curvature tensor. The comma is used to denote the ordinary derivative, $\kappa_{\mu,\nu}=\partial\kappa_\mu/\partial x_\nu$.

[3] F. London, Z. Physik 42, 375 (1927).

[4] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, 1923) p. 204.

We see at once that the right-hand side of (6) is a sum of a symmetrical and an antisymmetrical part. It is quite natural, therefore, to identify the second-rank tensor (except for a scale factor) $^*G_{\mu\nu}$ with the tensor $\mathcal{G}_{\mu\nu}$ introduced in (1). We shall therefore proceed by placing

$$^*G_{\mu\nu}=Q\mathcal{G}_{\mu\nu}, \tag{7}$$

where Q is a scale factor. On equating the symmetric and antisymmetric parts of (7) separately, we obtain

$$-2\mathfrak{F}_{\mu\nu}=QF_{\mu\nu}, \tag{8}$$

$$G_{\mu\nu}-(\kappa^\alpha{}_{,\alpha}-2\kappa_\alpha\kappa^\alpha)g_{\mu\nu}-2\kappa_\mu\kappa_\nu-(\kappa_{\mu,\nu}+\kappa_{\nu,\mu})=Q\lambda g_{\mu\nu}, \tag{9}$$

or

$$G_{\mu\nu}-2\kappa_\mu\kappa_\nu-(\kappa_{\mu,\nu}+\kappa_{\nu,\mu})=(\kappa^\alpha{}_{,\alpha}-2\kappa_\alpha\kappa^\alpha+Q\lambda)g_{\mu\nu}. \tag{10}$$

We shall now choose λ so that (10) will lead to the Klein-Gordon wave equation. We first place

$$\kappa^\alpha{}_{,\alpha}-2\kappa_\alpha\kappa^\alpha+Q\lambda=k', \tag{11}$$

so that we obtain for λ the expression

$$\lambda=(k'-\kappa^\alpha{}_{,\alpha}+2\kappa_\alpha\kappa^\alpha)/Q, \tag{12}$$

and (10) becomes

$$G_{\mu\nu}-(\kappa_{\mu,\nu}+\kappa_{\nu,\mu})-2\kappa_\mu\kappa_\nu=k'g_{\mu\nu}, \tag{13}$$

where k' is some universal constant, having the dimensions of the reciprocal of a length squared.

If we multiply both sides of (13) by $g^{\mu\nu}$ and make use of the index raising and lowering property of the metric tensor, we obtain

$$G-2\kappa^\nu{}_{,\nu}-2\kappa^\nu\kappa_\nu=4k', \tag{14}$$

where G is the Gaussian curvature at a point and is equal to $8\pi\rho_0(dt/ds)^2\Gamma$. The quantities ρ_0 and Γ are the proper density of matter at the point and the Newtonian gravitational constant, respectively.

We shall now show that (14) leads to the Klein-Gordon equation. We first rewrite it in the form

$$\kappa^\nu{}_{,\nu}+\kappa^\nu\kappa_\nu=\tfrac{1}{2}G-2k', \tag{15}$$

and note that κ_ν in this equation is defined except for a change of gauge. It follows, therefore, that (15) must still be valid if we replace κ_ν by $\kappa_\nu+\partial S/\partial x_\nu$, where S is the dimensionless function defined in (3). Since $\kappa^\nu{}_{,\nu}$ is equal to $\partial\kappa^\nu/\partial x_\nu$, we obtain from (15) the equation

$$\frac{\partial}{\partial x_\nu}\left[\kappa^\nu+\frac{\partial S}{\partial x^\nu}\right]+\left[\kappa^\nu+\frac{\partial S}{\partial x^\nu}\right]\left[\kappa_\nu+\frac{\partial S}{\partial x_\nu}\right]=M, \tag{16}$$

where we have placed M equal to $\frac{1}{2}G-2k'$. It is now easy to see that (16) is identical with

$$e^{-S}\left[\frac{\partial}{\partial x_\nu}+\kappa_\nu\right]\left[\kappa^\nu+\frac{\partial S}{\partial x^\nu}\right]e^S=M, \tag{17}$$

where the bracket on the left is to operate on everything that follows it. Multiplying through on the left by e^S and using the relation $(\partial S/\partial x_\nu)e^S=(\partial/\partial x_\nu)e^S$, we can write (17) as

$$\left(\frac{\partial}{\partial x_\nu}+\kappa_\nu\right)\left(\kappa^\nu+\frac{\partial}{\partial x^\nu}\right)e^S=Me^S. \tag{18}$$

We shall now introduce the electromagnetic potentials by means of the relation

$$\kappa_\nu=-iA_\nu\eta, \tag{19}$$

where η has the dimension of a reciprocal charge. We also substitute for S the quantity $-i\eta\chi$. If we now multiply both sides of (18) by $(e/c)^2$ and divide through by η, we obtain the result

$$\left(\frac{e}{c\eta i}\frac{\partial}{\partial x_\nu}-\frac{e}{c}A_\nu\right)\left(\frac{e}{c\eta i}\frac{\partial}{\partial x^\nu}-\frac{e}{c}A^\nu\right)e^{-i\eta\chi}=-\frac{e^2}{c^2\eta^2}Me^{-i\eta\chi}. \tag{20}$$

If we now define $\eta=(1/\alpha e)$, where α is the fine structure constant, we can introduce Planck's constant by means of the relation $\hbar=(e/c\eta)$, and (20) becomes

$$\left(\frac{\hbar}{i}\frac{\partial}{\partial x_\nu}-\frac{e}{c}A_\nu\right)\left(\frac{\hbar}{i}\frac{\partial}{\partial x^\nu}-\frac{e}{c}A^\nu\right)\exp\left(-i\frac{e}{c\hbar}\chi\right)$$
$$=\hbar^2(2k'-\tfrac{1}{2}G)\exp\left(-i\frac{e}{c\hbar}\chi\right). \tag{21}$$

It is clear that we may identify this with the Klein-Gordon equation provided we identify the coefficient of the exponential on the right-hand side of (21) with the rest mass of the charge distribution by means of the equation

$$\hbar^2(2k'-\tfrac{1}{2}G)=m^2c^2. \tag{22}$$

We shall see that this agrees with the result that we shall obtain in solving for the electrostatic case with gravitational effects neglected.

We shall return now to the determination of the function λ. From (5), (8), and (19) we obtain

$$QF_{\mu\nu}=2i\eta(\partial A_\mu/\partial x_\nu-\partial A_\nu/\partial x_\mu). \tag{23}$$

If we now place $Q=2i\eta$, we obtain

$$\lambda=i(k+\mu A_\alpha A^\alpha+\sigma A^\alpha{}_{,\alpha})\equiv i\Lambda, \tag{24}$$

where we have placed

$$\mu=\tfrac{1}{2}\eta;\quad k=-k'/\eta;\quad \sigma=-\tfrac{1}{2}i.$$

The constant k is a universal constant whose value is very much larger than the electromagnetic field intensities usually met with in the laboratory. (This quantity plays a role here similar to that of the maximum field intensity introduced in the Born-Infeld theory.) We see that λ is both symmetric and invariant.

With the choice (24) of λ, we shall now make use of (1) to set up a Lagrangian function for the electromagnetic field. We start by writing down the expression for the generalized volume element of our space. It has been shown by Eddington[5] that the simplest such volume element is given by the invariant,

$$dV=(-|\mathcal{G}_{\mu\nu}|)^{\frac{1}{2}}d\tau, \tag{25}$$

where $|\mathcal{G}_{\mu\nu}|$ is the determinant associated with the tensor $\mathcal{G}_{\mu\nu}$, and $d\tau=dx_1dx_2dx_3dx_4$.

It has been customary in the past to identify the components of the antisymmetric tensor $F_{\mu\nu}$ with the electromagnetic field strengths and to derive Maxwell's equations from a Lagrangian of the form

$$\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}(-g)^{\frac{1}{2}}d\tau,$$

where g is the determinant of the metric tensor.

Born and Infeld do not start from this form of the Lagrangian but instead use what is essentially the invariant integral (25) with λ taken equal to a constant. Since a Lagrangian function of this type is an irrational function of the field components, one is led to a set of nonlinear field equations.

In this paper we shall also start from the generalized volume element (25), but with λ defined by (24). If we identify the vector field A_μ with the electromagnetic potentials, then its curl [as defined by (23)] allows us to identify the components of $F_{\mu\nu}$ with the electromagnetic field intensities.

RATIONALIZING THE GENERALIZED VOLUME

If we neglect gravitational effects, we can choose a Galilean coordinate system and we obtain

$$|\mathcal{G}_{\mu\nu}|=\begin{vmatrix}-\lambda & f_{21} & f_{31} & f_{41}\\ -f_{21} & -\lambda & f_{32} & f_{42}\\ -f_{31} & -f_{32} & -\lambda & f_{43}\\ -f_{41} & -f_{42} & -f_{43} & \lambda\end{vmatrix}$$
$$=-\{\lambda^4+\lambda^2(f_{21}^2+f_{32}^2+f_{31}^2-f_{41}^2-f_{42}^2-f_{43}^2)-(f_{33}f_{41}-f_{31}f_{42}+f_{21}f_{43})^2\}\equiv-\mathcal{G}^2, \tag{26}$$

where the f_{ij} are components of the antisymmetric tensor $F_{\mu\nu}$.

If we now introduce the conventional notation for the potentials and the field strengths, we obtain

$$\mathcal{G}^2=\lambda^4+(\mathbf{B}^2-\mathbf{E}^2)\lambda^2-(\mathbf{B}\cdot\mathbf{E})^2, \tag{27}$$

where

$$\begin{gathered}\lambda=i[k+\mu(\mathbf{A}^2-\phi^2)+\sigma(\nabla\cdot\mathbf{A}+\dot{\phi}/c)],\\ \mathbf{E}=-\nabla\phi-(1/c)\partial\mathbf{A}/\partial t,\quad \mathbf{B}=\nabla\times\mathbf{A}.\end{gathered} \tag{28}$$

E and **B** are the electromagnetic field strengths, **A** is the vector potential, and ϕ is the scalar potential. The dot signifies differentiation with respect to the time.

Instead of using the form (27) for $\mathcal{G}^2$, we shall transform it to the expression

$$\mathcal{G}^2=(\mathbf{E}\times\mathbf{B})^2+[\lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]^2-\tfrac{1}{4}(\mathbf{B}^2+\mathbf{E}^2)^2. \tag{29}$$

[5] A. S. Eddington, reference 4, p. 234.

In order to obtain $\mathcal{G}$ as a rational function of the field strengths, we shall introduce certain auxiliary field quantities in the form of fourfold row and column matrices which we define as follows:

$$\Psi=\begin{Bmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{Bmatrix};\quad \Psi^+=(\psi_1\psi_2\psi_3\psi_4). \tag{30}$$

The product of these two quantities is defined in the usual way by means of the relation

$$\Psi^+\Psi=\psi_1^2+\psi_2^2+\psi_3^2+\psi_4^2. \tag{31}$$

For the time being we shall make no assumptions about the ψ's, which may either be constants or space-time functions.

We now multiply the expression (29) on the left by the row matrix in (30) and on the right by the column matrix in (30). If we introduce Λ instead of λ as defined by (24), we obtain

$$\Psi^+\mathcal{G}^2\Psi=\Psi^+\{(\mathbf{E}\times\mathbf{B})^2+[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]^2-\tfrac{1}{4}(\mathbf{B}^2+\mathbf{E}^2)^2\}\Psi. \tag{32}$$

We can factor this expression in a manner similar to that used by Dirac in setting up his relativistic wave equation. We introduce the Dirac matrices γ_1, γ_2, γ_3, γ_4 and $\gamma_5=\gamma_1\gamma_2\gamma_3\gamma_4$. If we let γ represent the first three matrices, then it can be seen that (32) factors into the two expressions

$$\begin{aligned}\mathcal{G}\Psi&=-\{\gamma\cdot(\mathbf{E}\times\mathbf{B})+\gamma_4[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]+\tfrac{1}{2}i\gamma_5(\mathbf{B}^2+\mathbf{E}^2)\}\Psi,\\ \Psi^+\mathcal{G}&=-\Psi^+\{\gamma\cdot(\mathbf{E}\times\mathbf{B})+\gamma_4[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]+\tfrac{1}{2}i\gamma_5(\mathbf{B}^2+\mathbf{E}^2)\}.\end{aligned} \tag{33}$$

We shall now take for our Lagrangian density the quantity $\mathcal{G}\Psi$ and write

$$\mathfrak{L}=-\{\gamma\cdot(\mathbf{E}\times\mathbf{B})+\gamma_4[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]+\tfrac{1}{2}i\gamma_5(\mathbf{B}^2+\mathbf{E}^2)\}\Psi. \tag{34}$$

We see that the Dirac matrices operate on Ψ in such a way as to replace the usual Lagrangian density by a fourfold Lagrangian density according to the following scheme:

$$\begin{aligned}\mathfrak{L}_1\equiv\mathcal{G}\psi_1&=-[(\mathbf{E}\times\mathbf{B})_x-i(\mathbf{E}\times\mathbf{B})_y]\psi_4-(\mathbf{E}\times\mathbf{B})_z\psi_3-[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_1-\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)\psi_3,\\ \mathfrak{L}_2\equiv\mathcal{G}\psi_2&=-[(\mathbf{E}\times\mathbf{B})_x+i(\mathbf{E}\times\mathbf{B})_y]\psi_3+(\mathbf{E}\times\mathbf{B})_z\psi_4-[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_2-\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)\psi_4,\\ \mathfrak{L}_3\equiv\mathcal{G}\psi_3&=-[(\mathbf{E}\times\mathbf{B})_x-i(\mathbf{E}\times\mathbf{B})_y]\psi_2-(\mathbf{E}\times\mathbf{B})_z\psi_1+[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_3+\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)\psi_1,\\ \mathfrak{L}_4\equiv\mathcal{G}\psi_4&=-[(\mathbf{E}\times\mathbf{B})_x+i(\mathbf{E}\times\mathbf{B})_y]\psi_1+(\mathbf{E}\times\mathbf{B})_z\psi_2+[-\Lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_4+\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)\psi_2.\end{aligned} \tag{35}$$

We may consider the ψ-quantities as additional degrees of freedom (additional fields of a nonelectromagnetic nature) that are introduced by the process of rationalization. These fields factor out when one passes from the linear to the quadratic Lagrangian density.

Since the use of a Lagrangian density of the form (35) will give rise to a fourfold set of field equations, it may be argued that this will result in blackbody radiation intensities four times larger than those predicted by the Planck formula and therefore in disagreement with observation. That this is not the case can easily be seen from the fact that all the ψ's are not of the same order of magnitude. In fact it can be shown that for weak fields, that is, for fields that are small compared to k or to Λ, the quantities ψ_1 and ψ_2 are small compared to ψ_3 and ψ_4. From this it follows that for weak fields the Lagrangian density defined above reduces to that obtained by Dirac.[2] To see this we rewrite (35) in the form

$$\begin{aligned}
\{\mathcal{G}+\lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)\}\psi_1 &= i[(\mathbf{E}\times\mathbf{B})_y+i(\mathbf{E}\times\mathbf{B})_x]\psi_4 \\
&\quad -[(\mathbf{E}\times\mathbf{B})_z+\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)]\psi_3, \\
\{\mathcal{G}+\lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)\}\psi_2 &= -[(\mathbf{E}\times\mathbf{B})_x+i(\mathbf{E}\times\mathbf{B})_y]\psi_3 \\
&\quad +[(\mathbf{E}\times\mathbf{B})_z-\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)]\psi_4, \\
\{\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)\}\psi_3 &= [i(\mathbf{E}\times\mathbf{B})_y-(\mathbf{E}\times\mathbf{B})_x]\psi_2 \\
&\quad -[(\mathbf{E}\times\mathbf{B})_z-\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)]\psi_1, \\
\{\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)\}\psi_4 &= -[(\mathbf{E}\times\mathbf{B})_x+i(\mathbf{E}\times\mathbf{B})_y]\psi_1 \\
&\quad +[(\mathbf{E}\times\mathbf{B})_z+\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)]\psi_2,
\end{aligned} \tag{36}$$

and note that for small $\mathbf{E}$ and $\mathbf{B}$ we may place $\mathcal{G}$ equal to λ^2. If we do this and solve for ψ_1 and ψ_2, we obtain

$$\begin{aligned}
\psi_1 &= \{[i(\mathbf{E}\times\mathbf{B})_y-(\mathbf{E}\times\mathbf{B})_x]\psi_4 \\
&\quad -[(\mathbf{E}\times\mathbf{B})_z+\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)]\psi_3\}/2\lambda^2, \\
\psi_2 &= \{-(\mathbf{E}\times\mathbf{B})_x+i(\mathbf{E}\times\mathbf{B})_y]\psi_3 \\
&\quad +[(\mathbf{E}\times\mathbf{B})_z-\tfrac{1}{2}(\mathbf{B}^2+\mathbf{E}^2)]\psi_4\}/2\lambda^2.
\end{aligned} \tag{37}$$

If we substitute these values of ψ_1 and ψ_2 into the last two equations of (36), we obtain the results

$$\begin{aligned}
[\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_4 &= [(\mathbf{E}\times\mathbf{B})^2-\tfrac{1}{4}(\mathbf{B}^2+\mathbf{E}^2)^2]\psi_4/2\lambda^2, \\
[\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_3 &= [(\mathbf{E}\times\mathbf{B})^2-\tfrac{1}{4}(\mathbf{B}^2-\mathbf{E}^2)^2]\psi_3/2\lambda^2.
\end{aligned} \tag{38}$$

If we now make use of the original definition of $\mathcal{G}^2$ as given in (29), we can substitute for the right-hand side of (38) and obtain

$$\begin{aligned}
[\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_4 &= [\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)] \\
&\quad \times[\mathcal{G}+\lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_4/2\lambda^2, \\
[\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_3 &= [\mathcal{G}-\lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)] \\
&\quad \times[\mathcal{G}+\lambda^2+\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2)]\psi_3/2\lambda^2.
\end{aligned} \tag{39}$$

From (39) we obtain at once the Lagrangian density for weak fields:

$$\mathcal{G}=-\Lambda^2-\tfrac{1}{2}(\mathbf{B}^2-\mathbf{E}^2). \tag{40}$$

For Λ different from zero this is the Lagrangian density used by Dirac, and for Λ equal to zero it reduces to the usual classical Lagrangian density. We see, then, that for the case of weak fields our theory gives the same result as that obtained with the classical theory.

If we now take (34) as our Lagrangian density, the Lagrangian becomes

$$L=\int \mathcal{L}d\tau, \tag{41}$$

and the variational principle becomes

$$\delta\int \mathcal{L}d\tau=0. \tag{42}$$

THE DERIVATION OF THE FIELD EQUATIONS

The variation of $\mathcal{L}$ with respect to the electromagnetic field leads to the equation

$$\begin{aligned}
\delta\mathcal{L} = -\{&\boldsymbol{\gamma}\cdot(\delta\mathbf{E}\times\mathbf{B})+\boldsymbol{\gamma}\cdot(\mathbf{E}\times\delta\mathbf{B}) \\
&+\gamma_4(-2\Lambda\delta\Lambda+\mathbf{B}\cdot\delta\mathbf{B}-\mathbf{E}\cdot\delta\mathbf{E}) \\
&+i\gamma_5(\mathbf{B}\cdot\delta\mathbf{B}+\mathbf{E}\cdot\delta\mathbf{E})\}\Psi.
\end{aligned} \tag{43}$$

If we now introduce the vector and scalar potentials from (28) and discard quantities that vanish at the integration limits, we find

$$\begin{aligned}
\delta\mathcal{L} = \{&-\delta\mathbf{A}\cdot[\boldsymbol{\gamma}\times(\nabla\times\mathbf{E})+\nabla\times(\boldsymbol{\gamma}\times\mathbf{E}) \\
&+\gamma_4\{-4\mu\mathbf{A}\Lambda+\nabla\times\mathbf{B}-(1/c)\partial\mathbf{E}/\partial t\} \\
&+i\gamma_5\{\nabla\times\mathbf{B}+(1/c)\partial\mathbf{E}/\partial t\}]-\delta\phi[\boldsymbol{\gamma}\cdot(\nabla\times\mathbf{B}) \\
&+\gamma_4(4\mu\phi\Lambda-\nabla\cdot\mathbf{E})+i\gamma_5\nabla\cdot\mathbf{E}]\}\Psi.
\end{aligned} \tag{44}$$

To obtain the field equations, we set the coefficient of $\delta\mathbf{A}$ and that of $\delta\phi$ separately equal to zero. After some trivial transformations, we obtain the equations of the electromagnetic field in the following form:

$$\nabla\times\mathbf{E}+(1/c)\partial\mathbf{B}/\partial t=0,\quad \nabla\cdot\mathbf{B}=0; \tag{45}$$

and

$$\begin{aligned}
&\nabla\times[(\gamma_4+i\gamma_5)\Psi\mathbf{B}+(\boldsymbol{\gamma}\times\mathbf{E})\Psi] \\
&\quad -(1/c)\frac{\partial}{\partial t}[(\gamma_4-i\gamma_5)\Psi\mathbf{E}+(\boldsymbol{\gamma}\times\mathbf{B})\Psi]=4\gamma_4\mu\Lambda\mathbf{A}\Psi, \\
&\nabla\cdot[(\gamma_4-i\gamma_5)\mathbf{E}+(\boldsymbol{\gamma}\times\mathbf{B})]\Psi=4\gamma_4\mu\Lambda\phi\Psi.
\end{aligned} \tag{46}$$

The first set is obtained from (28) and the second set from (42).

We now define the auxiliary field vectors

$$\begin{aligned}
\mathbf{D}&=(\gamma_4-i\gamma_5)\mathbf{E}\Psi+(\boldsymbol{\gamma}\times\mathbf{B})\Psi, \\
\mathbf{H}&=(\gamma_4+i\gamma_5)\mathbf{B}\Psi+(\boldsymbol{\gamma}\times\mathbf{E})\Psi.
\end{aligned} \tag{47}$$

In terms of these vectors the Lagrangian density takes on the simple form

$$\mathcal{L}=\tfrac{1}{2}(\mathbf{D}\cdot\mathbf{E}-\mathbf{H}\cdot\mathbf{B})+\gamma_4\Lambda^2\Psi; \tag{48}$$

and the field equations become

$$\begin{aligned}\nabla\times\mathbf{E}+(1/c)\partial\mathbf{B}/\partial t=0;\quad \nabla\cdot\mathbf{B}=0,\\ \nabla\times\mathbf{H}-(1/c)\partial\mathbf{D}/\partial t=4\gamma_4\mu\Lambda\mathbf{A}\Psi;\\ \nabla\cdot\mathbf{D}=4\gamma_4\mu\Lambda\phi\Psi.\end{aligned}\tag{49}$$

We may note that in the Born-Infeld theory the auxiliary fields that are introduced also appear defined in terms of $\mathbf{E}$ and $\mathbf{B}$ but in a very complicated and irrational form.

As an immediate consequence of our field equations, we may note that the charge and current densities appear defined in terms of the electromagnetic potentials. For the charge and current densities, we obtain the expressions

$$4\pi\rho=4\gamma_4\mu\Lambda\Psi\phi,\quad (4\pi/c)\rho\mathbf{v}=4\gamma_4\mu\Lambda\Psi\mathbf{A},\tag{50}$$

where $\mathbf{v}$ is the velocity of the charge distribution. If we divide the second of these equations by the first, we obtain

$$\mathbf{v}=(\mathbf{A}/\phi)c.\tag{51}$$

We see that this is precisely the result obtained from the retarded potentials of a moving point charge in the usual theory. The retarded scalar and vector potentials for a moving point charge in the usual theory are given by

$$\begin{aligned}\phi&=e[\mathbf{r}+(1/c)\mathbf{v}\cdot\mathbf{r}]^{-1}|_{t-r/c},\\ \mathbf{A}&=(e\mathbf{v}/c)[\mathbf{r}+(1/c)\mathbf{v}\cdot\mathbf{r}]^{-1}|_{t-r/c},\end{aligned}$$

and their ratio leads to (51).

It has been shown that for the case of weak fields the Lagrangian density (34) reduces to the usual one. Using the same procedure it is possible to show that the field Eqs. (46) for the case of weak fields reduce to the usual set of Maxwell-Lorentz equations for a charge current distribution.

SOLUTION FOR THE STATIC CASE

For the electrostatic case $\mathbf{B}$ vanishes and the last equation in (49) becomes

$$(\gamma_4-i\gamma_5)\Psi\nabla\cdot\mathbf{E}=4\gamma_4\mu\Lambda\Psi\phi.\tag{52}$$

From this equation we now obtain the four equations

$$\begin{aligned}(\nabla^2\phi)(\psi_1-\psi_3)&=-4\Lambda\mu\phi\psi_1,\\ (\nabla^2\phi)(\psi_2-\psi_4)&=-4\Lambda\mu\phi\psi_2,\\ (\nabla^2\phi)(\psi_1-\psi_3)&=4\Lambda\mu\phi\psi_3,\\ (\nabla^2\phi)(\psi_2-\psi_4)&=4\Lambda\mu\phi\psi_4,\end{aligned}\tag{53}$$

where we have replaced $\nabla\cdot\mathbf{E}$ by $\nabla^2\phi$.

If we now add the first of these equations to the third one and the second to the fourth one, we obtain two identical differential equations for the potential ϕ,

$$\nabla^2\phi=-2\Lambda\mu\phi.\tag{54}$$

We can solve this equation for the case of spherical symmetry by successive approximations. Using the first equation in (28), we place Λ equal to $(k-\mu\phi^2)$ and rewrite (54) in the form

$$(d^2/dr^2)(r\phi)=-2\mu(k-\phi^2)(r\phi).\tag{55}$$

Since k is very much larger than ϕ^2 for ordinary fields, we neglect ϕ^2 on the right-hand side of (55) to a first approximation and obtain for ϕ the solution

$$\phi=(a/r)\exp[-(2|k|\mu)^{\frac{1}{2}}r],\tag{56}$$

where $|k|$ stands for the absolute value of k, which is a negative constant, and a is a constant of integration.

If we substitute this back into (55), we obtain the equation

$$(d/dr)(r^2d\phi/dr)=-2(\Lambda\mu ar)\exp[-(2|k|\mu)^{\frac{1}{2}}r],$$

or

$$(d/dr)(r^2E_r)=-2(\Lambda\mu ar)\exp[-(2|k|\mu)^{\frac{1}{2}}r].$$

If we again neglect ϕ^2 with respect to k on the right-hand side of this equation, we obtain

$$(d/dr)(r^2E_r)=(2|k|\Lambda\mu ar)\exp[-(2|k|\mu)^{\frac{1}{2}}r]$$

If we integrate this from zero to some finite value of r, we obtain the solution for the electrostatic field

$$E_r=(e/r^2)\{1-[(2|k|\mu)^{\frac{1}{2}}r+1]\exp[-(2|k|\mu)^{\frac{1}{2}}r]\},\tag{57}$$

where the constant a has been placed equal to $-e$.

We see that for large values of r Eq. (57) reduces to the ordinary expression for the electrostatic field of a point charge. For small values of r (57) becomes

$$E_r=(e/r^2)\{1-[1+(2|k|\mu)^{\frac{1}{2}}r][1-(2|k|\mu)^{\frac{1}{2}}r]\},\tag{58}$$

and we see that the field takes on a finite value at the origin, namely,

$$E_{0r}=2|k|e\mu.\tag{59}$$

We shall now make use of the result for the electrostatic case to determine the total energy contained in the electrostatic field of a point charge at rest at the origin. Since the energy density is given by $E_r^2/8\pi$, we write for the total energy in the field the integral

$$U=\frac{1}{2}\int_0^\infty E_r^2r^2dr=\frac{e^2}{2}\int_0^\infty\{1-[(2|k|\mu)^{\frac{1}{2}}r+1]\times\exp[-(2|k|\mu)^{\frac{1}{2}}r]\}^2\frac{dr}{r^2}.\tag{60}$$

On carrying out the integration, we obtain

$$U=\frac{e^2}{2}\left\{-\frac{1}{r}-\frac{(2|k|\mu)^{\frac{1}{2}}}{2}\exp[-2(2|k|\mu)^{\frac{1}{2}}r]+\frac{2}{r}\exp[-(2|k|\mu)^{\frac{1}{2}}r]-\frac{1}{r}\exp[-2(2|k|\mu)^{\frac{1}{2}}r]\right\}\Bigg|_0^\infty.$$

At the upper limit the entire expression vanishes; and we obtain the value at the lower limit by expanding the exponentials to the first power in r. We thus find

$$U=\lim_{r\to 0}\frac{e^2}{2}\left\{\frac{1}{r}+\tfrac{1}{2}(2|k|\mu)^{\frac{1}{2}}-\frac{2}{r}[1-(2|k|\mu)^{\frac{1}{2}}r+\cdots]\right.$$
$$\left.+\frac{1}{r}[1-2(2|k|\mu)^{\frac{1}{2}}r+\cdots]\right\}.$$

On passing to the limit $r=0$, we finally obtain

$$U=\tfrac{1}{4}e^2(2|k|\mu)^{\frac{1}{2}}. \quad (61)$$

If we now equate the total energy in the field to the self energy mc^2 of the point charge, we obtain

$$\tfrac{1}{4}e^2(2|k|\mu)^{\frac{1}{2}}=mc^2. \quad (62)$$

From (62) we see that the classical radius of the electron, e^2/mc^2, enters into the theory quite naturally as the quantity $4/(2|k|\mu)^{\frac{1}{2}}$.

If we compare (62) with (22) we see that these two equations can be consistent only if

$$G=2(mc/\hbar)^2(32\alpha^2-1), \quad (63)$$

where α is the fine structure constant $\hbar c/e^2$. With this choice for G, we find that the constant $|k|$ has the value

$$|k|=8(mc/\hbar)^2\alpha^3 e. \quad (64)$$

We see from this that $|k|$ is indeed a very large constant, so that we are justified in neglecting ordinary fields compared with it.

If we substitute (64) into (59), we note that the electrostatic field at a point charge is just equal to $16(mc^2)^2/e^3$ or $2|k|/\alpha$.

THE HAMILTONIAN OF THE FIELD

To pass over to the Hamiltonian formalism, we introduce the momenta $\mathbf{p}_A$ and p_ϕ conjugate to the generalized coordinates $\mathbf{A}$, and ϕ, respectively, by means of the definitions:

$$p_A{}^\mu=\partial\mathcal{L}/\partial\dot{A}^\mu, \quad p_\phi=\partial\mathcal{L}/\partial\dot{\phi}. \quad (65)$$

We note that

$$\mathbf{p}_A=-(1/c)[(\gamma_4-i\gamma_5)\mathbf{E}+(\gamma\times\mathbf{B})]\Psi=-\mathbf{D}/c, \quad (66)$$

and that

$$p_\phi=-\gamma_4 i\Lambda\Psi/c=-\gamma_4\lambda\Psi/c. \quad (67)$$

We now define the Hamiltonian density as

$$\mathcal{H}=\mathbf{p}_A\cdot\partial\mathbf{A}/\partial t+p_\phi\cdot\dot{\phi}-\mathcal{L}. \quad (68)$$

If we make use of (48) for the Lagrangian density and place $\mathbf{D}=\mathbf{D}'\Psi$ and $\mathbf{H}=\mathbf{H}'\Psi$, the Hamiltonian density becomes

$$\mathcal{H}=\{\tfrac{1}{2}(\mathbf{D}'\cdot\mathbf{E}+\mathbf{H}'\cdot\mathbf{B})+\mathbf{D}'\cdot\nabla\phi+\gamma_4\lambda[\lambda+(\phi/c)]\}\Psi. \quad (69)$$

Since this Hamiltonian density contains no square roots or other irrational functions of the field variables, it should be possible to pass over to the quantum theory in the usual manner. The presence of the Dirac matrices in $\mathbf{D}'$ and $\mathbf{H}'$ in (69) should enable one to introduce the necessary spin into the theory.

DISCUSSION

There are a few points in connection with the theory developed above that are worth noting. The first point of interest concerns the manner in which the field equations and the Klein-Gordon equation enter into the theory. The former are obtained from a variational principle whereas the Klein-Gordon equation is introduced by means of a gauge transformation applied to a scalar equation. It would appear that the solution of the problem of the motion of an electron or a point charge involves the simultaneous solution of a set of field equations for the electromagnetic potentials and a set of wave equations. Since gravitational effects have been neglected, one may wonder whether a complete variational principle involving variation with respect to the gravitational as well as the electromagnetic potentials would not give both the field equations and the Klein-Gordon equation.

We have noted previously that the quantities ψ are not completely specified. In the present paper, we assumed them to be constants in dealing with the electromagnetic field equations. It is possible, however to allow the ψ's to enter the theory as space-time functions in which case the field equations would involve additional terms which depend on the derivatives of these quantities. In considering the nature of the Lagrangian density for weak fields, we do in fact tacitly assume that the ψ's are variable quantities such that for weak fields the first two components are small compared with the last two components.

We may finally note that the present theory does not suffer from the defect of having to introduce new arbitrary constants. The constants that do appear like k, G, and μ can all be expressed in terms of the fundamental constants of nature, e, m, h, and c.

In the summer of 1946, when the ideas developed in this paper first originated with the author, he had the pleasure of a number of stimulating discussions on the subject with Professor John A. Wheeler. It is also a pleasure to thank Peter G. Bergmann for suggesting some of the features of the arrangement of the material in the paper.