Derivation of Maxwell's equations from the gauge invariance of classical mechanics

Donald H. Kobe
Department of Physics, North Texas State University, Denton, Texas 76203
(Received 12 September 1979; accepted 1 November 1979)

The Lagrangian for a single classical charged particle is made form invariant under the addition of a total time derivative by adding an interaction Lagrangian which involves compensating fields. The compensating fields are the vector and scalar potentials of the electromagnetic field which couple to the current and charge densities, respectively. To insure form invariance of the Lagrangian, the vector and scalar potentials must undergo the usual gauge transformations of electromagnetism. The electric and magnetic fields, which are gauge invariant, are obtained by examining the equation of motion for the charged particle. Faraday's law and the condition that there are no magnetic monopoles are obtained from the expressions for the electric and magnetic fields in terms of the potentials. The simplest possible gauge-invariant Lagrangian which is quadratic in the electric and magnetic fields is constructed. From the principle of least action Gauss' law and the Ampère–Maxwell law are obtained.

I. INTRODUCTION

Gauge invariance in classical electrodynamics is not emphasized in most textbooks. Perhaps this is because Maxwell's equations and the Lorentz force involve the fields, not the potentials, and so are obviously gauge invariant. These equations and Newton's second law or its relativistic generalization are all that is necessary to treat classical electrodynamic problems. It is only when classical electrodynamics is formulated in terms of the Lagrangian[1] that the electromagnetic potentials play an essential role. However, the equivalence of Lagrangians that differ only by a gauge transformation on the potentials[2] is often either not mentioned[3] or relegated to the problems.[4,5]

This situation contrasts sharply with the current research emphasis in quantum field theory on gauge transformations.[6] Gauge theories are now thought to be the key to understanding all the basic interactions of physics, and there is optimism that the long sought goal of a unified field theory may be attained through a gauge theory.[7] The vast number of possible Lagrangians describing the fundamental interactions can be sharply reduced by using only those interactions that arise by making the Lagrangian form invariant under local gauge transformations.[8]

The derivation of Maxwell's equations by the method of gauge theory was first made by Weyl.[9] By considering the Lagrangian for a Dirac electron to be form invariant under local gauge transformations on the spinor wave functions he was led to introduce the four-vector potential as a compensating field with certain transformation properties. The electromagnetic field strength tensor was defined in terms of the potentials in the usual way in order to eliminate the arbitrary gauge function. He constructed the simplest gauge invariant scalar Lagrangian for the electromagnetic field, from which the two dynamical Maxwell equations, Gauss' law and the Ampère–Maxwell law, follow from the principle of least action. The two kinematical Maxwell equations, Faraday's law and the condition of no magnetic monopoles, can be obtained directly from the definition of the electromagnetic field strength tensor in terms of the vector potential. This method was again used by Schwinger.[10] Yang and Mills[11] applied the same approach to local rotations in isospin space to obtain the field that now bears their name.

In a previous paper[12] a simplified derivation of Maxwell's equations from the local gauge invariance of quantum mechanics was given. Instead of using the Lagrangian of quantum field theory, the role of gauge invariance for the Schrödinger equation was discussed. The form invariance of the Schrödinger equation under local gauge transformations led to the introduction of the vector and scalar potentials as compensating fields.[13] The two dynamical Maxwell equations were derived from energy conservation considerations. The two kinematical Maxwell equations followed from the definitions of the electric and magnetic fields in terms of the potentials.

The derivation of Maxwell's equations as a gauge theory is not limited to the use of quantum field theories, but can even be done by considering the form invariance of the Lagrangian for a classical charged particle. In this paper Maxwell's equations are derived by considering classical mechanics from the gauge theory point of view. To make the Lagrangian for a classical charged particle form invariant under the addition of a total time derivative a vector and scalar potential can be introduced which couple to the charge current and density, respectively. Form invariance leads to the gauge transformation properties of the potentials. When the principle of least action is applied to the charged particle in the new field Newton's second law is obtained with the Lorentz force. The usual expressions for the electric and magnetic fields in terms of the potentials are obtained from the Lorentz force. As usual, the kinematical Maxwell equations are obtained directly from the form of the electric and magnetic fields in terms of the potentials. The dynamical Maxwell equations are obtained from the principle of least action by constructing the simplest gauge-invariant scalar Lagrangian for the field.

The presentation used here, as well as in all gauge theory approaches, to derive Maxwell's equations turns around the traditional order in which electromagnetism is presented.[1-4] Instead of considering the vector and scalar potentials as auxiliary fields which are only useful for calculation, they are introduced as compensating fields to preserve the form of the Lagrangian of a classical particle under the addition of a total time derivative. In this paper the *gauge principle* states that the Lagrangian should have the same form when a total time derivative is added to it, and is thus the same as the *principle of form invariance* under gauge transfor-

 0002-9505/80/050348-06$00.60 © 1980 American Association of Physics Teachers

mation. The method of constructing gauge theories is perhaps strange to students encountering it for the first time. However, when they see the method applied to the familiar subject of electromagnetism, they may have less difficulty in understanding the more abstract Yang–Mills field later. The presentation here, which uses the electric and magnetic fields instead of the electromagnetic field strength tensor, could be used in a senior or graduate course in electromagnetism or classical mechanics. It shows the deep insight that can be gained into the former from the latter.

In Sec. II, the form of the Lagrangian for a single classical particle is shown to be changed under the addition of a total time derivative. In Sec. III, compensating fields, which are the electromagnetic vector and scalar potentials, are introduced, which couple to the charge current and density, respectively. The new Lagrangian is now form invariant under the addition of a total time derivative. Newton's second law with the Lorentz force is obtained in Sec. IV from the principle of least action with the new Lagrangian. In Sec. V Maxwell's equations are derived. The two kinematical equations, Faraday's law and the condition of no magnetic monopoles, are obtained from the definition of the electric and magnetic fields in terms of the potentials. The two dynamical equations, Gauss' law and the Ampère–Maxwell law, are obtained by constructing the simplest gauge-invariant quadratic Lagrangian, and using the principle of least action. The conclusion is given in Sec. VI. An appendix gives the covariant formulation.

II. TRANSFORMATION OF THE LAGRANGIAN

To illustrate the approach used here, we shall consider a hypothetical physicist who knows classical mechanics but does not know anything about electromagnetism. We shall follow his thinking, which when guided by intuition and an esthetic sense leads him to discover Maxwell's equations on the basis of the symmetry of the Lagrangian. Conceivably such a discovery of Maxwell's equations could have been made at any time after Lagrange[14] (1736–1813) by someone sufficiently ingenious. The fact that Maxwell's equations were actually developed as a synthesis of experimental laws is, however, not surprising. The approach used here is a strictly modern one, which emphasizes the symmetry of the Lagrangian.[15] Classical physics has traditionally emphasized the equations of motion, and considered the symmetry of the Lagrangian under various operations as an interesting, but basically irrelevant, by-product of the equations of motion. On the other hand, modern physics has emphasized symmetry principles as a guide to discover new dynamical equations. However, to treat electromagnetism from the modern point of view should help the student understand the approach of gauge theories in a familiar realm.

Our hypothetical physicist considers a single classical particle with mass m and displacement $\mathbf{r} = \mathbf{r}(t)$ at time t, which has an external potential energy $U(\mathbf{r})$ not of electromagnetic origin. The Lagrangian for this particle is

$$L_0 = (1/2)m\dot{\mathbf{r}}^2 - U(\mathbf{r}), \tag{2.1}$$

where $\dot{\mathbf{r}}$ is the velocity of the particle. The action S_0 associated with this Lagrangian is

$$S_0 = \int_1^2 dt\, L_0, \tag{2.2}$$

where the integral is over the time interval t_1 to t_2. By varying the action S_0 with respect to the displacement $\mathbf{r}(t)$, keeping the end points fixed,

$$\delta S_0 = 0, \tag{2.3}$$

he obtains Newton's second law

$$m\ddot{\mathbf{r}} = -\nabla U(\mathbf{r}), \tag{2.4}$$

where the right-hand side of Eq. (2.4) is the force acting on the particle. Equation (2.3) is the principle of least (or stationary) action.[14]

Our physicist discovers that adding a total time derivative to the Lagrangian does not change the equation of motion.[16] The time derivative of an arbitrary differentiable function $(q/c)\Lambda(\mathbf{r},t)$, where q and c are parameters introduced for convenience now which are determined later, can be added to Eq. (2.1) to obtain the new Lagrangian L_0'

$$L_0' = L_0 + \frac{q}{c}\frac{d\Lambda}{dt}. \tag{2.5}$$

The action S_0' is calculated by replacing L_0 in Eq. (2.2) by Eq. (2.5), which gives

$$S_0' = S_0 + (q/c)[\Lambda(\mathbf{r}_2,t_2) - \Lambda(\mathbf{r}_1,t_1)], \tag{2.6}$$

where $\mathbf{r}_2 = \mathbf{r}(t_2)$ and similarly for $\mathbf{r}_1$. Since the variation of the terms in the square brackets gives zero because the end points are fixed, the use of Eq. (2.6) in Hamilton's principle in Eq. (2.3) also gives Newton's second law in Eq. (2.4). He therefore discovers that the Lagrangian in Eq. (2.5) is equivalent to the original Lagrangian in Eq. (2.1).

Our physicist speculates that in addition to its mass m, the particle may have another quantity q, which he calls "charge", associated with it.[17] The charge of a body is responsible for its being attracted or repelled by other charged bodies. For the point particle the density of charge is

$$\rho(\mathbf{x},t) = q\delta[\mathbf{x} - \mathbf{r}(t)], \tag{2.7}$$

where $\mathbf{x}$ is an arbitrary point is space. The charge density vanishes everywhere except at $\mathbf{r}(t)$, where it is infinity. However, its integral over all space is q. The quantity $\delta(\mathbf{x} - \mathbf{r})$ has the property that it is zero except for $\mathbf{x} = \mathbf{r}$, but when integrated over all space it is one. The charge current is a vector quantity,

$$\mathbf{J}(\mathbf{x},t) = q\dot{\mathbf{r}}\delta[\mathbf{x} - \mathbf{r}(t)], \tag{2.8}$$

which also vanishes unless $\mathbf{x} = \mathbf{r}$. The delta function $\delta(\mathbf{x} - \mathbf{r})$ was invented by Dirac,[18] but in principle it could have been invented earlier by our ingenious physicist.

The Lagrangian in Eq. (2.5) can be rewritten in terms of the charge density and current. Since the displacement $\mathbf{r}$ is a function of time, the total time derivative in Eq. (2.5) is

$$\frac{d\Lambda(\mathbf{r},t)}{dt} = \frac{\partial\Lambda(\mathbf{r},t)}{\partial t} + \dot{\mathbf{r}}\cdot\frac{\partial\Lambda(\mathbf{r},t)}{\partial\mathbf{r}}. \tag{2.9}$$

When Eq. (2.9) is substituted into Eq. (2.5) and Eqs. (2.7) and (2.8) are used, our physicist can write the new Lagrangian in the form

$$L_0' = L_0 + \frac{1}{c}\int d^3x \left(\mathbf{J}(\mathbf{x},t)\cdot\nabla\Lambda(\mathbf{x},t) + c\rho(\mathbf{x},t)\frac{1}{c}\frac{\partial\Lambda(\mathbf{x},t)}{\partial t}\right), \tag{2.10}$$

where $\nabla = \partial/\partial \mathbf{x}$ is the gradient with respect to the field point $\mathbf{x}$, and the integral is over all space.

A relation between the current $\mathbf{J}$ and the density ρ can be obtained by performing an integration by parts in Eq. (2.10),

$$L_0' = L_0 - \frac{1}{c}\int d^3x\left(\nabla\cdot\mathbf{J} + \frac{\partial\rho}{\partial t}\right)\Lambda + \frac{d}{dt}\frac{1}{c}\int d^3x\,\rho\Lambda, \quad (2.11)$$

where it is assumed that the current at infinity is zero. In order for Eq. (2.11) to reduce to Eq. (2.5) when Eq. (2.7) is substituted into it, the first integral on the right-hand side of Eq. (2.11) must vanish. Since Λ is an arbitrary function of space and time, the only way that the integral can vanish identically is if

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J} = 0, \quad (2.12)$$

which is the equation of continuity for the charge. If Eqs. (2.7) and (2.8) are substituted into the equation of continuity, it is satisfied exactly. If Eq. (2.12) is integrated over all space, we find that the charge q is a constant in time.

The Lagrangian L_0' in Eq. (2.10) is completely equivalent to the original Lagrangian L_0 is Eq. (2.1). However, our physicist has a sense that something is incomplete. He feels that all equivalent equations in physics should have the same form. He speculates that the integral in Eq. (2.10) is indicative of a force field[19] in nature which can couple to electric charge.

III. FORM INVARIANCE OF THE LAGRANGIAN

Our physicist feels that equivalent equations in physics should have the same form. This feeling is basically esthetic, but he postulates it as a principle. Equation (2.10) is certainly not the same form as Eq. (2.1) unless Λ is a constant. In general, Λ is an arbitrary function of space and time. In order to impose this *principle of form invariance* of the Lagrangian under a total time derivative, he is led to a modification of Eq. (2.10).

In order to construct a new Lagrangian that is form invariant under the addition of a total time derivative, our physicist must add a term to the Lagrangian in Eq. (2.1) which is similar in form to the integral in Eq. (2.10). Since $\nabla\Lambda$ is a vector, he speculates that there must be a vector field $\mathbf{A}(\mathbf{x},t)$ which couples to the charge current $\mathbf{J}(\mathbf{x},t)$ of the particle. Since $\partial\Lambda/\partial t$ is a scalar, he speculates that there must also be a scalar field $A_0(\mathbf{x},t)$ which couples to the charge density $\rho(\mathbf{x},t)$ of the particle. The field $\mathbf{A}$ he calls the "vector potential" and the field A_0 he calls the "scalar potential." He thus adds a new interaction Lagrangian L_i to the particle Lagrangian L_0 in Eq. (2.1) to obtain a new Lagrangian

$$L_{o,i} = L_o + \frac{1}{c}\int d^3x(\mathbf{J}\cdot\mathbf{A} - c\rho A_0), \quad (3.1)$$

where the last term is the interaction L_i. The new Lagrangian in Eq. (3.1) describes a single charged particle interacting with a new field, called the "electromagnetic field" by our physicist, which couples to the particle's charge.

To obtain an equivalent Lagrangian, our physicist adds a total time derivative to Eq. (3.1),

$$L_{0,i}' = L_{0,i} + \frac{q}{c}\frac{d\Lambda}{dt}. \quad (3.2)$$

The total time derivative in Eq. (3.2) is equal to the integral in Eq. (2.10). When Eq. (3.1) is substituted into Eq. (3.2) and the total time derivative is written as the integral in Eq. (2.10), our physicist obtains

$$L_{0,i}' = L_0 + \frac{1}{c}\int d^3x(\mathbf{J}\cdot\mathbf{A}' - c\rho A_0'), \quad (3.3)$$

which has the same form as Eq. (3.1). Equation (3.3) is obtained if the new fields $\mathbf{A}'$ and A_0' are defined as

$$\mathbf{A}' = \mathbf{A} + \nabla\Lambda \quad (3.4)$$

and

$$A_0' = A_0 - \frac{1}{c}\frac{\partial\Lambda}{\partial t}. \quad (3.5)$$

The vector and scalar potentials in Eq. (3.1) absorb the additional terms in Eq. (2.10) and become transformed. He calls Eqs. (3.4) and (3.5) "gauge transformations." (Perhaps the term "gradient transformations," which the Russians use,[20] would have been more appropriate.) The fields $\mathbf{A}$ and A_0 are in this sense "compensating fields"[13] which are introduced to maintain the form invariance of the Lagrangian under the addition of a total time derivative. The addition of a total time derivative to the Lagrangian could be called a "gauge transformation on the Lagrangian" since it induces gauge transformations on the potentials. The equations of motion for the charged particle in this new field can now be obtained from the principle of least action.

IV. EQUATION OF MOTION FOR THE CHARGED PARTICLE

The equations of motion for the charged particle can now be obtained by our physicist using the principle of least action. The equation of motion of the particle in the new field should give a means for detecting the new field, since it should modify the motion of the particle if it is present. There should be an additional force on a charged particle if it is in the new electromagnetic field. However, our physicist is puzzled by the apparent lack of uniqueness of the compensating fields in Eqs. (3.4) and (3.5). They may not be the physical fields.

The principle of least action can be used to obtain the equation of motion for the particle. The new action for the charged particle is

$$S_{0,i} = \int_1^2 dt\, L_{0,i}, \quad (4.1)$$

where the integral is from t_1 to t_2 and the Lagrangian $L_{0,i}$ in Eq. (3.1) is used. Our physicist varies Eq. (4.1) with respect to $\mathbf{r}(t)$ keeping the end point fixed. When he sets the variation equal to zero he obtains the equation of motion

$$m\ddot{\mathbf{r}} = -\nabla U + q\left(-\nabla A_0 - \frac{\partial\mathbf{A}}{\partial(ct)}\right) + \left(\frac{q}{c}\right)\dot{\mathbf{r}}\times(\nabla\times\mathbf{A}), \quad (4.2)$$

which is Newton's second law with additional forces coming from the electromagnetic field.

The question arises in our physicist's mind as to whether the equation of motion in Eq. (4.2) is unique. After all the Lagrangian $L'_{0,i}$ in Eq. (3.3) is equivalent to $L_{0,i}$. When he varies $S'_{0,i}$, obtained by replacing $L_{0,i}$ in Eq. (4.1) with $L'_{0,i}$, he obtains an equation of motion with the same form as Eq. (4.2),

$$m\ddot{\mathbf{r}} = -\nabla U + q\left(-\nabla A'_0 - \frac{\partial \mathbf{A}'}{\partial(ct)}\right) + \left(\frac{q}{c}\right)\dot{\mathbf{r}} \times (\nabla \times \mathbf{A}'). \quad (4.3)$$

The only difference between the Lagrangian $L_{0,i}$ in Eq. (3.1) and the Lagrangian $L'_{0,i}$ in Eq. (3.3) is that $\mathbf{A}$ and A_0 have been replaced by $\mathbf{A}'$ and A'_0 in Eqs. (3.4) and (3.5), respectively. Our physicist is a little puzzled at first because it looks as if Eq. (4.3) is different than Eq. (4.2), even though they have the same form. Then he looks at Eqs. (3.4) and (3.5), and realizes that Eqs. (4.2) and (4.3) are exactly the same equation.

The fields $\mathbf{A}$ and A_0 introduced in Eq. (3.1) cannot be the measureable physical fields, since by Eqs. (3.4) and (3.5) they are not unique. Our physicist defines a new field $\mathbf{B}$,

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (4.4)$$

which he calls the "magnetic field." It does not make any difference in Eq. (4.4) whether $\mathbf{A}$ or $\mathbf{A}'$ in Eq. (3.4) is used, since the curl of the gradient of Λ is zero. Likewise, he defines another field $\mathbf{E}$,

$$\mathbf{E} = -\nabla A_0 - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \quad (4.5)$$

which he calls the "electric field." It does not make any difference in Eq. (4.5) whether $\mathbf{A}$ and A_0 or $\mathbf{A}'$ and A'_0 in Eqs. (3.4) and (3.5), respectively, are used, since the space and time derivatives of Λ commute. The magnetic field $\mathbf{B}$ and the electric field $\mathbf{E}$ are the physical fields since they are invariant under the gauge transformations in Eqs. (3.4) and (3.5).

The equation of motion of the particle in Eq. (4.2) or Eq. (4.3) can now be written in a manifestly gauge invariant way,

$$m\ddot{\mathbf{r}} = -\nabla U + q\mathbf{E} + \left(\frac{q}{c}\right)\dot{\mathbf{r}} \times \mathbf{B}. \quad (4.6)$$

Equation (4.6) is Newton's second law with the Lorentz force, although the force due to the magnetic fields was first given by Heaviside.[21] It can be used as an operational definition to determine whether an electric or magnetic field is present by measuring the acceleration of a test body of mass m and charge q.

V. DERIVATION OF MAXWELL'S EQUATIONS

Having obtained the electric and magnetic fields, our physicist wonders what equations they satisfy. In order to produce them and to calculate their values he must know their equations. He finds that they satisfy four equations, of which two are kinematical and two are dynamical. The two kinematical equations are obtained from the equations for the fields in terms of the potentials. The two dynamical equations he obtains from the principle of least action, using a Lagrangian for the electromagnetic field.

Our physicist obtains the two kinematical equations satisfied by $\mathbf{E}$ and $\mathbf{B}$ directly from their definitions in terms of the vector and scalar potentials. Taking the curl of Eq. (4.5) for $\mathbf{E}$ and using Eq. (4.4), he obtains

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad (5.1)$$

which is Faraday's law. Taking the divergence of Eq. (4.4) for $\mathbf{B}$, he obtains

$$\nabla \cdot \mathbf{B} = 0, \quad (5.2)$$

which is the statement that there are no magnetic monopoles.

However, he realizes that Eqs. (5.1) and (5.2) are not sufficient to determine completely $\mathbf{E}$ and $\mathbf{B}$. He conjectures that the principle of least action can also be used for the fields $\mathbf{E}$ and $\mathbf{B}$ if a proper Lagrangian L_{em} for the electromagnetic field can be found.[1] The Lagrangian is a scalar quantity, and the simplest way to construct a quadratic scalar quantity from the vectors $\mathbf{E}$ and $\mathbf{B}$ is to take their dot products. He therefore constructs a Lagrangian

$$L_{em} = \int d^3x(\alpha E^2 + \beta B^2 + \gamma \mathbf{E} \cdot \mathbf{B}), \quad (5.3)$$

where α, β, and γ are arbitrary constants to be chosen later.[22] He considers adding terms like $\mathbf{A}^2$ and A_0^2 to the integrand in Eq. (5.3), but realizes that the form invariance of the total Lagrangian under the addition of a total time derivative would be destroyed by these terms.[23]

The total Lagrangian L is the sum of the Lagrangian in Eq. (3.1) of the particle interacting with the field and the Lagrangian of the electromagnetic field in Eq. (5.3),

$$L = L_{em} + L_0 + L_i. \quad (5.4)$$

Using the potentials $\mathbf{A}$ and A_0 as generalized coordinates, our physicist can obtain the dynamical equations for the field from the principle of least action. The action S calculated from the Lagrangian L in Eq. (5.4) is

$$S = \int_1^2 dt\, L, \quad (5.5)$$

where the time interval goes from t_1 to t_2. Equations (4.4) and (4.5) are substituted into L_{em} for $\mathbf{B}$ and $\mathbf{E}$, respectively, before the variation is performed with respect to $\mathbf{A}$ and A_0.

When our physicist varies Eq. (5.5) with respect to A_0 he obtains

$$2\alpha\nabla \cdot \mathbf{E} = \rho - \gamma\nabla \cdot \mathbf{B}. \quad (5.6)$$

From Eq. (5.2) the last term on the right-hand side of Eq. (5.6) vanishes. Since the units of charge q are arbitrary, he chooses the constant α to be

$$\alpha = 1/8\pi. \quad (5.7)$$

Therefore Eq. (5.6) becomes

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \quad (5.8)$$

which is Gauss' law in Gaussian units.

When he varies the action in Eq. (5.5) with respect to $\mathbf{A}$ he obtains

$$\frac{2\alpha}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{1}{c}\mathbf{J} + 2\beta\nabla \times \mathbf{B} = -\gamma\left(\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}\right). \quad (5.9)$$

The right-hand side of Eq. (5.9) vanishes because of Eq. (5.1). Substituting Eq. (5.7) into Eq. (5.9), he obtains

$$-8\pi\beta\nabla\times\mathbf{B}=\frac{4\pi}{c}\mathbf{J}+\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}, \qquad (5.10)$$

which involves the unknown constant β.

The constant β must now be determined. In the special case that $\mathbf{J}=0$, the curl of Eq. (5.10) gives

$$-8\pi\beta\nabla\times(\nabla\times\mathbf{B})=\frac{1}{c}\frac{\partial}{\partial t}\nabla\times\mathbf{E}. \qquad (5.11)$$

When he uses the identity for the curl of a curl of a vector in Eq. (5.11) along with Eqs. (5.1) and (5.2), he obtains

$$\nabla^2\mathbf{B}+(8\pi\beta c^2)^{-1}\frac{\partial^2\mathbf{B}}{\partial t^2}=0. \qquad (5.12)$$

If $\beta<0$, Eq. (5.12) is the equation for a magnetic wave traveling with a speed of $(8\pi|\beta|)^{1/2}c$, where c in Eq. (2.5) is taken as positive. On the other hand, if $\beta>0$ a wave solution to Eq. (5.12) does not exist, and it is not clear how to interpret the constant $8\pi\beta c^2$. Moreover, our physicist is indeed excited about the possibility of magnetic waves. Are there electric waves also? Taking the curl of Eq. (5.1) and using Eq. (5.10), he obtains in the special case of $\rho=0$ and $\mathbf{J}=0$,

$$\nabla^2\mathbf{E}+(8\pi\beta c^2)^{-1}\frac{\partial^2\mathbf{E}}{\partial t^2}=0. \qquad (5.13)$$

If $\beta<0$ there are also electric waves, which propagate with the same speed $(8\pi|\beta|)^{1/2}c$. Because of Eq. (5.1) the electric and magnetic waves are related and form what he calls an "electromagnetic wave."

Since c in Eq. (2.5) is an arbitrary positive constant with the dimensions of velocity, he chooses c to be the velocity of propagation of the electromagnetic wave. The constant β then becomes

$$\beta=-1/8\pi. \qquad (5.14)$$

Therefore Eq. (5.10) becomes

$$\nabla\times\mathbf{B}=\frac{4\pi}{c}\mathbf{J}+\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}, \qquad (5.15)$$

which is the Ampère–Maxwell law in Gaussian units. By taking the divergence of Eq. (5.15) and using Gauss's law in Eq. (5.8), he obtains the equation of continuity in Eq. (2.12).

Our physicist has now obtained all four of Maxwell's equations in Eq. (5.1), (5.2), (5.8), and (5.15). The parameters α and β in the Lagrangian $L_{\rm em}$ in Eq. (5.3) have been obtained in Eqs. (5.7) and (5.14), respectively. However, the parameter γ in Eq. (5.3) has not been determined. Since it does not occur in the dynamical equations, it can be taken without loss of generality to be zero, $\gamma=0$.

In retrospect our physicist can understand why the choice $\gamma=0$ can be made. From Eq. (4.6) he sees that $\mathbf{E}$ is a vector and that $\mathbf{B}$ is a pseudovector, since the classical equation of motion should be invariant with respect to parity. Thus the first two terms on the right-hand side of Eq. (5.3) are scalars, while the last term in a pseudoscalar. If the Lagrangian is a pure scalar, then $\gamma=0$.

Our physicist naturally asks whether the dynamical equations obtained for the field would have been different had he used the total Lagrangian L',

$$L'=L_{\rm em}+L_0+L_i', \qquad (5.16)$$

instead of L in Eq. (5.4) to calculate the action S' from Eq. (5.5). The Lagrangian L' in Eq. (5.16) has the same form as the Lagrangian L in Eq. (5.4). The new action S' would be varied with respect to the generalized coordinates $\mathbf{A}'$ and A_0'. Since the equations obtained by varying S with respect to $\mathbf{A}$ and A_0 involve only $\mathbf{E}$ and $\mathbf{B}$, the same results would be obtained by varying S' with respect to $\mathbf{A}'$ and A_0'. Therefore the same dynamical equations in Eqs. (5.8) and (5.15) would be obtained.[24] The formulation is therefore completely gauge invariant because of the form invariance of the Lagrangian.

VI. CONCLUSION

Our physicist who did not know anything about electromagnetism at the beginning of this paper has now derived Maxwell's equations and the Lorentz force guided by his esthetic sense of the form invariance of the Lagrangian under the addition of a total time derivative. In principle our physicist could have lived any time after Lagrange.[14] From a realistic point of view this approach could only be expected in the late twenties, when the symmetry properties of the Lagrangian began to be emphasized with the advent of quantum field theory.[25] It was 1929 before the approach of gauge invariance was used to derive Maxwell's equations from a Lagrangian for a relativistic electron.[9]

The approach used in this paper illustrates the approach used in any gauge theory.[26] The emphasis is on the symmetry of the Lagrangian under a basic transformation. A compensating field is introduced to insure the form invariance of the Lagrangian under the transformation. The compensating field is a potential from which the physical field is obtained by differentiation in such a way that the arbitrary gauge function is eliminated. The equations of motion for the field are obtained by constructing the simplest nontrivial Lagrangian, and using the principle of least action.

The Lagrangian for a classical nonrelativistic charged particle is used here only to illustrate the method of approach. Maxwell's equations are invariant under Lorentz transformations, while classical nonrelativistic mechanics is invariant under Galilean transformations. When our physicist realizes this incompatibility, he may be led to formulate relativistic mechanics, as Einstein did.[27] Nevertheless, the approach used here shows the deep connection between mechanics and electromagnetism, while illustrating the approach of modern gauge theories.

APPENDIX: COVARIANT FORMULATION

The approach of this paper can be formulated more concisely by using the four-vector potential.[28] If $\mathcal{L}_0$ is the Lagrangian density for the particle, then the equation of motion for the particle follows by setting the variation of the action

$$S_0=\int d^4x\,\mathcal{L}_0 \qquad (\mathrm{A}1)$$

equal to zero. A four divergence can be added to $\mathcal{L}_0$ without changing the equation of motion. The new Lagrangian $\mathcal{L}_0'$ is[29]

$$\mathcal{L}_0'=\mathcal{L}_0+\partial_\mu(J^\mu\Lambda) \qquad (\mathrm{A}2)$$

where $\partial_\mu \equiv \partial/\partial x^\mu (\mu = 0,1,2,3)$, J^μ is a current four-vector associated with the particle, and Λ is an arbitrary function of x^μ. The summation convention is used in Eq. (A2) and subsequent equations. Equation (A2) can be rewritten as

$$\mathcal{L}_0' = \mathcal{L}_0 + \Lambda\partial_\mu J^\mu + J^\mu \partial_\mu \Lambda, \tag{A3}$$

so that the form of $\mathcal{L}_0'$ is different from $\mathcal{L}_0$.

A new four vector A_μ, called the "four potential," which couples to J^μ can be introduced as a "compensating field" to absorb the last term in Eq. (A3). If we add the new interaction term $\mathcal{L}_i = -J^\mu A_\mu$ to Eq. (A1), we obtain

$$\mathcal{L}_{0,i} = \mathcal{L}_0 - J^\mu A_\mu. \tag{A4}$$

Then when the four divergence in Eq. (A2) is added to Eq. (A4), we obtain

$$\mathcal{L}_{0,i}' = \mathcal{L}_0' - J^\mu A_\mu. \tag{A5}$$

If Eq. (A3) is substituted into Eq. (A5), we obtain

$$\mathcal{L}_{0,i}' = \mathcal{L}_0 - J^\mu A_\mu' + \Lambda \partial_\mu J^\mu. \tag{A6}$$

The new four potential A_μ' is defined as

$$A_\mu' = A_\mu - \partial_\mu \Lambda. \tag{A7}$$

The principle of form invariance is satisfied if

$$\partial_\mu J^\mu = 0, \tag{A8}$$

i.e., the charge is conserved. Then Eq. (A6) becomes

$$\mathcal{L}_{0,i}' = \mathcal{L}_0 - J^\mu A_\mu', \tag{A9}$$

which has exactly the same form as Eq. (A4). Thus the principle of form invariance leads not only to the interaction term, but to charge conservation in Eq. (A8) as well.

The electromagnetic field strength tensor $F_{\mu\nu}$ can be defined as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{A10}$$

Equation (A10) can also be used with the potential in Eq. (A7) since the arbitrary gauge term cancels. From the form of Eq. (A10) we obtain

$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0, \tag{A11}$$

where $(\alpha,\beta,\gamma) = (0,1,2,3)$ with no number repeated. When $(\alpha,\beta,\gamma) = (1,2,3)$ in Eq. (A11), the condition of no magnetic monopoles is obtained. When $\alpha = 0$ in Eq. (A11) Faraday's law is obtained.

The simplest scalar Lagrangian which can be constructed for the electromagnetic field is

$$\mathcal{L}_{\mathrm{em}} = -\frac{1}{16\pi} F_{\mu\nu}F^{\mu\nu}, \tag{A12}$$

where the indices on $F_{\mu\nu}$ are raised using the metric tensor $g^{\alpha\beta} = \mathrm{diag}(1,-1,-1,-1)$. The constant in Eq. (A12) is chosen to determine the units of charge to be Gaussian units. When Eq. (A12) is added to Eq. (A4), and the principle of least action is applied to the field we obtain

$$\partial_\mu F^{\mu\nu} = (4\pi/c)J^\nu, \tag{A13}$$

which is Gauss' law for $\nu = 0$ and the Ampère–Maxwell law for $\nu = (1,2,3)$.

[1] A. Sommerfeld, *Electrodynamics* (Academic, New York, 1952), pp. 269–273. [He ascribes the principle of least action for the electromagnetic field to K. Schwarzschild, Göttingen Nachr. Ges. Wiss., Math.-Phys. Kl. **1903**, 126, 132, 245 (1903).

[2] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 3rd ed. (Addison-Wesley, Reading, MA, 1971), pp. 49–50.

[3] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950), Chap. 11.

[4] J. D. Jackson, *Classical Electrodynamics*, 2nd ed., (Wiley, New York, 1975), p. 613, Prob. 12.2.

[5] D. E. Soper, *Classical Field Theory* (Wiley, New York, 1976), p. 100, Prob. 2.

[6] J. C. Taylor, *Gauge Theories of Weak Interactions* (Cambridge University, Cambridge, 1976).

[7] See, e.g., S. Weinberg, Am. Sci. **65**, 171 (1977); S. Weinberg, Phys. Today **30**, 42 (April, 1977); S. Weinberg, Sci. Am. **231**, 50 (July, 1974).

[8] For a recent review of gauge theories, see L. O'Raifeartaigh, Rep. Progr. Phys. **42**, 159 (1979).

[9] H. Weyl, Z. Phys. **56**, 330 (1929).

[10] J. Schwinger, Phys. Rev. **91**, 713 (1953). See, especially pp. 724–725.

[11] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[12] D. H. Kobe, Am. J. Phys. **46**, 342 (1978).

[13] F. A. Kaempffer, *Concepts in Quantum Mechanics* (Academic, New York, 1965), Secs. 21 and 22.

[14] R. Dugas, *A History of Mechanics* (Central, New York, 1955). (See Part III, Chap. 11 for a discussion of Lagrange's contributions to mechanics, and Part III, Chap. 5 on the principle of least action. Hamilton's contributions to the principle of least action are given in Part IV, Chap. 6.)

[15] E. P. Wigner, *Symmetries and Reflections* (MIT, Cambridge, MA, 1970), Chaps. 1–4.

[16] L. D. Landau and E. M. Lifshitz, *Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1969) pp. 2–4.

[17] D. Roller and D. H. D. Roller, *The Development of the Concept of the Electric Charge* (Harvard University, Cambridge, MA, 1954).

[18] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th revised ed. (Oxford University, London, 1967), pp. 58–61.

[19] M. B. Hesse, *Forces and Fields* (Philosophical Library, New York, 1962), Chap. VIII.

[20] See, e.g., A. I. Akhiezer and V. B. Berestetskii, *Quantum Electrodynamics* (Interscience, New York, 1965), p. 155.

[21] E. Whittaker, *A History of the Theories of Aether and Electricity*, Vol. 1 (Nelson, London, 1951), pp. 310, 395–396.

[22] The last term in the integrand in Eq. (5.3) is not a scalar, but a pseudoscalar, and so γ can be taken to be zero. However, we shall show that the value of γ does not affect the dynamical equations.

[23] These terms are included in the Proca Lagrangian for a massive vector field. A Proca, J. Phys. Radium **7**, 347 (1936). See also Ref. 4, p. 598.

[24] This gauge invariant procedure was not followed by E. A. Power and T. Thirunamachandran, Am. J. Phys. **46**, 370 (1978). After making a gauge transformation they vary with respect to the old potentials. The equation they obtain is correct but not form invariant. They obtain only the transverse part of the Ampere–Maxwell law and do not obtain Gauss' law.

[25] For a survey of the history of quantum field theory, see S. Weinberg,, Daedalus **106** (4), 17 (1977).

[26] R. Utiyama, Phys. Rev. **101**, 1597 (1956).

[27] A. Einstein, Ann. Phys. (Leipzig) **17**, 891 (1905). This paper has been translated and reprinted in *The Principle of Relativity* (Dover, New York, 1952), pp. 35–65.

[28] See Ref. 4, pp. 547–552, 595–597.

[29] In this appendix we choose units such that $c = 1$.