International Journal of Theoretical Physics, Vol. 21, No. 8/9, 1982

Gauge Invariance and the Dirac Equation[1]

Donald H. Kobe

Department of Physics, North Texas State University, Denton, Texas 76203

Received September 10, 1981

The gauge invariance of the Dirac equation is reviewed and gauge-invariant operators are defined. The Hamiltonian is shown to be gauge dependent, and an energy operator is defined which is gauge invariant. Gauge-invariant operators corresponding to observables are shown to satisfy generalized Ehrenfest theorems. The time rate of change of the expectation value of the energy operator is equal to the expectation value of the power operator. The virial theorem is proved for a relativistic electron in a time-varying electromagnetic field. The conventional approach to probability amplitudes, using the eigenstates of the unperturbed Hamiltonian, is shown in general to be gauge dependent. A gauge-invariant procedure for probability amplitudes is given, in which eigenstates of the energy operator are used. The two methods are compared by applying them to an electron in a zero electromagnetic field in an arbitrary gauge.

1. INTRODUCTION

The source of these complications is that although our total Hamiltonian is gauge-invariant, we split it into two parts, H_0 and V, which are not separately gauge-invariant, and then assume that V is small. It is not very satisfactory to have a whole perturbation technique based on the assumption of the smallness of a quantity... which is not gauge invariant (Dirac, 1965).

Dirac is here referring to quantum electrodynamics, where the total Hamiltonian is for both matter and quantized radiation. The total Hamiltonian is also the total energy of the system, which is a gauge-invariant quantity (up to a constant). The complications to which he refers are those connected with a gauge-invariant form of quantum electrodynamics he developed (Dirac, 1955) which "is more complicated as many new terms appear in the equations... (whose) physical significance is not clear" (Dirac, 1965). Dirac (1950, 1952) has also explored gauge invariance in other

[1]Presented at the Dirac Symposium, Loyola University, New Orleans, May 1981.

0020-7748/82/0800-0685$03.00/0

papers. Without mentioning its relationship with the gauge problem, Dirac (1958) proved the independence of the Kramers–Heisenberg formula from the form of the interaction in the electric dipole approximation ($\mathbf{E}\cdot\mathbf{r}$ or $\mathbf{A}\cdot\mathbf{p}$ plus A^2, constants omitted).

Dirac is so dissatisfied with the divergences of quantum electrodynamics and the renormalization scheme for handling them, that he has explored the possibility of giving up gauge invariance if that would solve the difficulties. He believes the reason that quantum electrodynamics gives infinities is that the wrong classical theory is being quantized. A new classical electron theory was proposed (Dirac, 1951) in which the superfluous variables related to the gauge freedom of the electromagnetic potentials acquire physical significance and describe electric charges. The simplest relativistically invariant condition on the potentials $A_\mu A^\mu = k^2$ is used, where k is a universal physical constant (which turns out to be m/e). This assumption is the only one he needs to get classical electrons appearing in the theory. Although his program was not completed, since the classical theory was not quantized, it shows how strongly Dirac abhors the divergences in quantum electrodynamics (see also Dirac, 1970).

This paper does not deal with the truly fundamental question of divergences or even the question of gauge transformations in quantum electrodynamics. Instead it deals with the more modest problem of an electron in a hydrogen atom in the presence of an external classical electromagnetic field. The ideas of manifest gauge invariance of the Dirac equation are reviewed. In particular, the gauge invariance of the probability amplitude for finding the electron in an energy eigenstate is considered, to which the quotation at the beginning is also appropriate.

Yang (1976) developed a manifestly gauge-invariant formulation of nonrelativistic quantum mechanics (Kobe and Smirl, 1978). He observed that the Hamiltonian for a particle in an *external* classical time-dependent electromagnetic field is gauge dependent, and cannot be the energy. The energy operator, which is manifestly gauge invariant, is defined to be the sum of the kinetic and potential energies. The usual expansion coefficients in quantum mechanics are in general gauge dependent, and cannot be interpreted as probability amplitudes. Yang (1976) defines gauge-invariant probability amplitudes by using eigenstates of the energy operator (Cohen-Tannoudji et al., 1977).

Most of the ideas of gauge invariance in the nonrelativistic case can be applied in a straightforward manner to the relativistic case. The role of gauge invariance in the reduction of the Dirac equation for an electron in a time-dependent electromagnetic field to the corresponding Pauli equation using the Foldy–Wouthuysen transformation has been discussed in a previous paper (Kobe and Yang, 1980) and is not reviewed here (Yang, 1982).

In Section 2 the form invariance of the Dirac equation under gauge transformations is reviewed. The gauge invariance of operators is discussed in Section 3. Ehrenfest's theorem is applied to several operators of physical interest, including the energy operator, in Section 4. The conventional approach to probability amplitudes is shown to be in general gauge dependent in Section 5. Gauge-invariant probability amplitudes are discussed in Section 6. In Section 7 the conventional and gauge-invariant formulations are applied to the case of an electron in a zero electromagnetic field. A comparison of the conventional and gauge-invariant probability amplitudes is made in Section 8, and it is shown under what conditions they agree. Finally, the conclusions are given in Section 9.

2. GAUGE INVARIANCE OF THE DIRAC EQUATION

The gauge invariance of the Dirac equation (Rose, 1961) is proved in the same way as the gauge invariance of the Schrödinger equation (Bohm, 1951). However, to review the concepts and establish the notation, the gauge invariance of the Dirac equation is proved here.

The Dirac Hamiltonian for a single particle of charge q and mass m in an external classical electromagnetic radiation field characterized by the four-potential $A^\mu = (A^0, A^1, A^2, A^3) = (A^0, -A_1, -A_2, -A_3)$ is

$$H(A^\mu) = c\boldsymbol{\alpha}\cdot(\mathbf{p} - q\mathbf{A}/c) + \beta mc^2 + V + qA^0 \tag{1}$$

where $\mathbf{p} = -i\hbar\nabla$ is the momentum operator and $\mathbf{A} = (A^1, A^2, A^3)$ (Dirac, 1928a, 1928b; Bjorken and Drell, 1964). The external potential energy of the particle is V. For example, for an electron in a hydrogen atom, the potential energy V is the Coulomb potential between the proton and the electron. Potential energy is a nonrelativistic concept, so the presence of V destroys the Lorentz invariance of the theory. If the potential energy V were zero, however, the theory would be Lorentz invariant (Sakurai, 1967). The matrices $\boldsymbol{\alpha} = (\alpha^1, \alpha^2, \alpha^3)$ and β satisfy the anticommutation relations

$$\{\alpha^i, \alpha^j\} = 2\delta^{ij}, \qquad \{\alpha^i, \beta\} = 0 \tag{2}$$

and $\beta^2 = 1$. The Dirac equation in its Hamiltonian form is (Feynman, 1962)

$$H(A^\mu)\psi = i\hbar\,\partial\psi/\partial t \tag{3}$$

which has the same form as the Schrödinger equation.

A gauge transformation on the spinor wave function is (Rose, 1961)

$$\psi' = \exp(iq\Lambda/\hbar c)\psi \tag{4}$$

where $\Lambda = \Lambda(x)$ is a differentiable function of the space-time coordinates $x = x^\mu = (x^0, x^1, x^2, x^3)$ and $x^0 = ct$. Under the transformation in equation (4) the Dirac equation in equation (3) becomes (Rose, 1961)

$$H(A'^\mu)\psi' = i\hbar\,\partial\psi'/\partial t \tag{5}$$

which has the same form as equation (3). The new potentials in equation (5) are

$$A'_\mu = A_\mu - \partial_\mu\Lambda \tag{6}$$

where $\partial_\mu = \partial/\partial x^\mu$, which is the usual gauge transformation of electromagnetism. The term "gauge invariant" applied to the Dirac equation means its form invariance under gauge transformations on both the wave function and the potentials as shown in equation (5).

The electromagnetic field strength tensor $F_{\mu\nu}$ is obtained from the potential A_μ by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{7}$$

This equation is invariant under the gauge transformation in equation (6) on the potential.

3. GAUGE INVARIANCE OF OPERATORS

For an operator to correspond to an observable, it must be Hermitian (Dirac, 1958). It must also be gauge invariant, so that the same expectation value is obtained in all gauges (Kobe and Smirl, 1978; Kobe and Yang, 1980). For an operator $\theta(A^\mu)$ which may depend on the potential, the gauge invariance of expectation values is

$$\langle\psi|\theta(A^\mu)\psi\rangle = \langle\psi'|\theta(A'^\mu)\psi'\rangle \tag{8}$$

i.e., the same value is obtained in the new gauge when the potentials in the operator are replaced by the new potentials and the expectation value is taken with respect to the new wave function. Since the wave function ψ' is defined by equation (4), we can write the left-hand side of equation (8) as

$$\langle\psi|\theta(A^\mu)\psi\rangle = \langle\psi'|\theta'(A^\mu)\psi'\rangle \tag{9}$$

where the operator θ' is the unitarily transformed operator

$$\theta'(A^\mu) = \exp(iq\Lambda/\hbar c)\theta(A^\mu)\exp(-iq\Lambda/\hbar c) \tag{10}$$

We shall call the unitary transformation in equation (10) a "gauge transformation on the operator." Comparing equations (8) and (9) with each other, we see that the operator must satisfy

$$\theta'(A^\mu) = \theta(A'^\mu) \tag{11}$$

in order to have a gauge-invariant expectation value. In other words, a gauge transformation on the operator induces a gauge transformation on the potentials on which the operator depends. Equation (11) says that the operator θ corresponding to an observable must be *form invariant* under gauge transformations (Kobe and Yang, 1980).

As an example of a gauge-invariant operator consider the kinetic four-momentum operator $\pi_\mu = p_\mu + qA_\mu/c$ where $p_\mu = -i\hbar\,\partial/\partial x^\mu = -i\hbar\,\partial_\mu$. Under a gauge transformation it transforms as

$$(\pi_\mu)' = (p_\mu + qA_\mu/c)' = p_\mu + qA'_\mu/c = \pi'_\mu \tag{12}$$

Equation (12) satisfies equation (11) so the kinetic four-momentum is a gauge-invariant operator. The canonical four-momentum operator p_μ alone is not a gauge-invariant operator, since its expectation value depends on the gauge or phase factor of the wave function (Cohen-Tannoudji et al., 1977).

The Hamiltonian in equation (1) is not a gauge-invariant operator, since

$$H'(A^\mu) = H(A'^\mu) + q\,\partial_0\Lambda \tag{13}$$

Because the scalar potential A^0 is not transformed under the unitary transformation, the Hamiltonian is not form invariant under gauge transformations. Its transformation properties are, however, such that the Dirac equation in equation (3) is form invariant.

Since the scalar potential causes the Hamiltonian to be gauge dependent, we shall define an energy operator (Kobe and Yang, 1980)

$$\mathcal{E}(A_i) = c\alpha^i(p_i + qA_i/c) + \beta mc^2 + V \tag{14}$$

where $A_i = -A^i$ $(i = 1,2,3)$ and a sum over the repeated index i from 1 to 3 is implied. The energy operator in equation (14) is the Hamiltonian $H(A^\mu)$ from which the scalar potential term qA^0 has been subtracted. This operator

is gauge invariant,

$$\mathcal{E}'(A_i) = \mathcal{E}(A_i') \tag{15}$$

which satisfies equation (11). We shall see in the next section that the operator in equation (14) is indeed the energy operator, since the time rate of change of its expectation value is equal to the power supplied by the external field.

4. EHRENFEST'S THEOREM

In this section Ehrenfest's theorem and some of its generalizations are presented. Ehrenfest's theorem (Yang, 1976) is important because it shows the correspondence between quantum and classical theory. It also aids in the interpretation of quantum theory. In Ehrenfest's theorem, the time rate of change of the expectation value of an operator corresponding to an observable is shown to be equal to the expectation value of another operator corresponding to another observable. The operators considered here in Ehrenfest's theorem are the displacement, kinetic momentum, total angular momentum, and energy. The first three operators have been considered by Feynman (1962) in terms of their Heisenberg equations of motion. The virial theorem is also proved for a particle in time-varying electromagnetic fields.

Consider first the displacement operator $\mathbf{r}$. The time rate of change of its expectation value is

$$d\langle\psi|\mathbf{r}\psi\rangle/dt = \langle\psi|\mathbf{v}\psi\rangle \tag{16}$$

which gives the velocity operator $\mathbf{v}$. If the Dirac equation in equation (3) is used for the time derivative of the wave functions in equation (16), the velocity operator is

$$\mathbf{v} = c\boldsymbol{\alpha} \tag{17}$$

which is a well-known result (Sakurai, 1967).

Consider the kinetic three-momentum operator $\pi^i = i\hbar\partial^i - qA^i/c$ $(i = 1,2,3)$. The time rate of change of its expectation value is

$$d\langle\psi|\boldsymbol{\pi}\psi\rangle/dt = \langle\psi|\mathbf{F}\psi\rangle \tag{18}$$

where $\mathbf{F}$ is the quantum mechanical force operator. If the Dirac equation in equation (3) is used for the time rate of change of the wave function in

equation (18), the force operator is

$$\mathbf{F} = -\nabla V + q\mathbf{E} + (q/c)\mathbf{v}\times\mathbf{B} \tag{19}$$

where $\mathbf{v}$ is the velocity operator in equation (17) which commutes with the magnetic induction vector $\mathbf{B}$, and $\mathbf{E}$ is the electric field.

The time rate of change of the average total angular momentum $\mathbf{J}$ is

$$d\langle\psi|\mathbf{J}\psi\rangle/dt = \langle\psi|\mathbf{r}\times\mathbf{F}\psi\rangle \tag{20}$$

The operator $\mathbf{r}\times\mathbf{F}$, where $\mathbf{F}$ is the force operator in equation (19), is the quantum mechanical torque operator. The total angular momentum operator $\mathbf{J}$ is

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \tag{21}$$

where $\mathbf{L} = \mathbf{r}\times\boldsymbol{\pi}$ is the orbital angular momentum and $\boldsymbol{\pi}$ is the kinetic three-momentum in equation (12). The intrinsic angular momentum or spin $\mathbf{S}$ is defined as $S_i = (-i\hbar/2)\alpha_j\alpha_k$, where (ijk) is a cyclic permutation of $(1,2,3)$.

The time rate of change of the average energy is

$$d\langle\psi|\mathcal{E}(A_i)\psi\rangle/dt = \langle\psi|P\psi\rangle \tag{22}$$

where $\mathcal{E}(A_i)$ is the energy operator in equation (14). The quantum mechanical power operator P is defined as (Yang, 1976; Kobe and Yang, 1980)

$$P = q\mathbf{v}\cdot\mathbf{E} \tag{23}$$

where $\mathbf{v}$ is the velocity operator in equation (17). Equation (22) verifies the claim made at the end of Section 3 that $\mathcal{E}(A_i)$ is the energy operator because the time rate of change of its expectation value is equal to the average power supplied to the electron by the external electromagnetic field.

The virial theorem for a relativistic electron in a time-dependent electromagnetic field can also be considered along with the other generalizations of Ehrenfest's theorem. The virial theorem for an electron in a time-dependent electromagnetic field has the same form as for static fields (Rose and Welton, 1952; Schectman and Good, 1956). The operator considered is the virial $\mathbf{r}\cdot\boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the kinetic three-momentum in equation (12). The time rate of change of the expectation value of the virial is

$$d\langle\psi|\mathbf{r}\cdot\boldsymbol{\pi}\psi\rangle/dt = \langle\psi|(\mathbf{r}\cdot\mathbf{F} + \mathbf{v}\cdot\boldsymbol{\pi})\psi\rangle \tag{24}$$

The operator $\mathbf{v}\cdot\boldsymbol{\pi} = c\boldsymbol{\alpha}\cdot\boldsymbol{\pi}$, which in the classical case is twice the kinetic energy. In the relativistic case there is of course a distinction between the velocity operator $c\boldsymbol{\alpha}$ in equation (17) and $\boldsymbol{\pi}/m$.

5. CONVENTIONAL PROCEDURE FOR PROBABILITY AMPLITUDES

All quantities that can in principle be measured must be gauge invariant. This principle means that the probability that a system is in a given state must also be gauge invariant. However, the conventional approach to probability amplitudes using the unperturbed Hamiltonian H_0 and its eigenfunctions gives gauge-dependent amplitudes in general. In this section the conventional procedure is reviewed and criticized from the standpoint of gauge invariance (Yang, 1982).

The conventional procedure for time-dependent problems is to write the Dirac equation in equation (3) as

$$\{H_0 - q\boldsymbol{\alpha}\cdot\mathbf{A} + qA^0\}\psi = i\hbar\,\partial\psi/\partial t \tag{25}$$

The unperturbed Dirac Hamiltonian,

$$H_0 = c\boldsymbol{\alpha}\cdot\mathbf{p} + \beta mc^2 + V \tag{26}$$

satisfies the eigenvalue equation

$$H_0\phi_n = E_n\phi_n \tag{27}$$

where the eigenfunctions ϕ_n form a complete set.

In the conventional approach the amplitude

$$a_n(t) = \langle\phi_n|\psi(t)\rangle \tag{28}$$

is taken as the probability for finding the system in the state ϕ_n at time t. Unless the electromagnetic field (and also the potentials) are zero at time t, equation (28) is not a probability amplitude for finding the system in an energy eigenstate, since H_0 in equation (26) is not the same as the energy operator in equation (14).

The equation satisfied by the amplitude in equation (28) can be obtained by taking its time derivative, using equation (25) and the completeness of states $\{\phi_n\}$ to give

$$i\hbar\dot{a}_n - E_n a_n = \sum_m \langle\phi_n|(-q\boldsymbol{\alpha}\cdot\mathbf{A} + \bar{q}A^0)\phi_m\rangle a_m \tag{29}$$

The exact solution to this equation with the specified initial conditions gives the amplitude in equation (28).

A gauge transformation on the Dirac equation gives equation (5). If someone follows the previous procedure in the new gauge, the amplitude for finding the system in the state ϕ_n is

$$a'_n(t) = \langle \phi_n | \psi'(t) \rangle \tag{30}$$

From equation (4) this amplitude can be written as

$$a'_n(t) = \langle \phi_n | \exp(iq\Lambda/\hbar c)\psi(t) \rangle \tag{31}$$

Equation (31) is not the same as equation (28) because Λ is in general a function of both the space and time. Therefore, in general

$$|a_n(t)|^2 \neq |a'_n(t)|^2 \tag{32}$$

The amplitude a_n cannot in general be interpreted as a probability amplitude for finding the system in an energy eigenstate because its absolute value squared is not gauge invariant. A probability must of course be gauge invariant. If Λ in equation (31) depended on the time only, the equality would hold in equation (32), but this case is not general.

The equation which equation (30) satisfies is obtained by taking its time derivative, which gives in a manner similar to equation (29)

$$i\hbar \dot{a}'_n - E_n a'_n = \sum_m \langle \phi_n | (-q\boldsymbol{\alpha} \cdot \mathbf{A}' + qA'^0)\phi_m \rangle a'_m \tag{33}$$

If this equation were solved exactly with the appropriate initial conditions, we would obtain equation (30).

The objection may be raised (Epstein, 1979) that it is not ϕ_n that should be used in equation (30), but the state

$$\phi'_n = \exp(iq\Lambda/\hbar c)\phi_n \tag{34}$$

Equation (34) is an eigenstate of H'_0,

$$H'_0 \phi'_n = E_n \phi'_n \tag{35}$$

where H'_0 is the unitarily transformed H_0. The unitary transformation in equation (10) on H_0 gives

$$H'_0 = c\boldsymbol{\alpha} \cdot (\mathbf{p} - q\nabla\Lambda/c) + \beta mc^2 + V \tag{36}$$

If the eigenstate of H'_0 in equation (34) were used in equation (30) instead of

eigenstates of H_0, the inner product would be

$$a_n(t) = \langle \phi_n'(t) | \psi'(t) \rangle \tag{37}$$

The phase factors on both functions in equation (37) cancel, which gives a_n in equation (28). However, the operator H_0 is not *form invariant* under a unitary transformation, since H_0' in equation (36) does not have the same form as H_0 in equation (26). Therefore the operator H_0 is not gauge invariant in the sense discussed in Section 3, because it involves the gauge-dependent canonical momentum **p**.

Another way of looking at the problem is to consider two people who are each solving the same Dirac equation in a different gauge. If they each follow the *same* conventional procedure they will get different results for the probability that the system is in the state ϕ_n as equation (32) shows. If they communicate with each other, they can obtain the gauge function which connects their potentials and wave functions. Then the first could tell the second to transform his unperturbed Hamiltonian so that the second would agree with the first and get $|a_n(t)|^2$ for the probability that the system is in the state ϕ_n. On the other hand, the second could tell the first to transform *his* unperturbed Hamiltonian, so that the first would agree with the second and get $|a_n'(t)|^2$ for the probability that the system is in the state ϕ_n. Who would be correct? Neither would be, since in general neither $|a_n|^2$ nor $|a_n'|^2$ is the probability that the system is in an energy eigenstate.

In the next section, a manifestly gauge-invariant procedure is given for obtaining the probability that the system is in an energy eigenstate.

6. GAUGE INVARIANCE OF PROBABILITIES

Because of the gauge dependence of the conventional procedure discussed in the last section, it is important to obtain a manifestly gauge-invariant procedure for obtaining probability amplitudes (Yang, 1976; Kobe and Smirl, 1978). The reason the amplitudes in equations (28) and (30) are not probability amplitudes is that H_0 is not a gauge-invariant operator. In order to have a manifestly gauge-invariant procedure, only gauge-invariant operators should be used (Schwinger, 1951).

If we want to know the probability that the electron is in an energy eigenstate, it is necessary to use the eigenstates of the energy operator $\mathcal{E}(A_i)$ in equation (14). The energy operator satisfies the eigenvalue equation (Kobe and Yang, 1980)

$$\mathcal{E}(A_i)\psi_n = \varepsilon_n \psi_n \tag{38}$$

where ψ_n is the energy eigenstate and ε_n is the energy eigenvalue. Since the vector potential A_i depends on the time, the eigenstates ψ_n and eigenvalues ε_n also depend on the time as a parameter.

The probability amplitude at the time t that the electron is in an eigenstate of the energy operator is

$$c_n(t) = \langle \psi_n(t) | \psi(t) \rangle \tag{39}$$

By differentiating equation (39), using the Dirac equation in equation (3), and using the completeness of the states $\{\psi_n\}$, we can show that the amplitude in equation (39) satisfies the equation

$$i\hbar \dot{c}_n - \varepsilon_n c_n = \sum_m \langle \psi_n | c\pi_0 \psi_m \rangle c_m \tag{40}$$

The operator $\pi_0 = -i\hbar\partial_0 + qA_0/c$ is the zero component of the kinetic four-momentum in equation (12). The exact solution of equation (40) with the specified initial conditions gives equation (39).

Equations (38)–(40) are all manifestly gauge invariant. They retain the same form if A^μ, ψ, and ψ_n are replaced by A'^μ, ψ', and ψ_n', respectively, where ψ_n' is defined as

$$\psi_n' = \exp(iq\Lambda/\hbar c)\psi_n \tag{41}$$

which is the same transformation as for ψ in equation (4). The eigenvalue ε_n in equation (38) and the amplitude in equation (39) are unchanged under the gauge transformations, and are thus gauge invariant. The matrix element in equation (40) is also gauge invariant

$$\langle \psi_n | \pi_0 \psi_m \rangle = \langle \psi_n' | \pi_0' \psi_m' \rangle \tag{42}$$

where π_0' is given in equation (12).

The amplitude in equation (39) is unchanged under gauge transformations because both ψ_n and ψ transform in the same way, so

$$c_n = \langle \psi_n' | \psi' \rangle \tag{43}$$

Since ψ_n' and ψ_n are gauge equivalent, they describe the same state. An eigenstate of the energy operator is denoted by $[\psi_n]$, the class of all wave functions that are gauge equivalent to ψ_n. In other words, $\psi_n' \in [\psi_n]$, if $\psi_n' = \exp(iq\Lambda/hc)\psi_n$, for some real function Λ. The probability that the

electron is in the state $[\psi_n]$ of the energy operator at the time t is thus

$$P_n = |c_n(t)|^2 \tag{44}$$

where $c_n(t)$ is given in equations (39) and (43). The gauge-invariant procedure gives the same probability in all gauges, since it is based on using gauge-invariant operators.

7. DIRAC EQUATION WITH A ZERO ELECTROMAGNETIC FIELD

As an example of the conventional procedure and the gauge-invariant procedure, we shall apply each to the Dirac equation where the electromagnetic field is zero. A zero electromagnetic field can be described by a nonzero potential which is the four-gradient of a scalar function. If there is no electromagnetic field, we know intuitively that there should be no transitions between states. The conventional approach predicts transitions between unperturbed states, unless the potentials are chosen to be zero. The gauge-invariant procedure predicts no transitions in any gauge.

If the electromagnetic field strength tensor in equation (7) is zero

$$F_{\mu\nu} = 0 \tag{45}$$

the potential A_μ must be of the form

$$A_\mu = \partial_\mu \Lambda \tag{46}$$

If the gauge transformation in equation (6) is made using equation (46) as the old potential, the new potential is

$$A'_\mu = 0 \tag{47}$$

The conventional amplitude a_n in equation (28) satisfies equation (29). With the potential in equation (46), equation (29) becomes

$$i\hbar \dot{a}_n - E_n a_n = \sum_m \langle \phi_n | \left(q\alpha^i \partial_i \Lambda + q\, \partial_0 \Lambda \right) \phi_m \rangle a_m \tag{48}$$

The quantity in parentheses on the right-hand side of equation (48) is a "pseudointeraction," since there is no electromagnetic field present. However, this pseudointeraction couples different states ϕ_n. Because of this coupling, we shall show that $|a_n(t)|^2$ depends on the time. Since there is no

electromagnetic field, the probability that the system is in a given state cannot depend on the time.

If we use the zero potential in equation (47) in equation (33) for the amplitude a'_n, we obtain

$$i\hbar\dot{a}'_n - E_n a'_n = 0 \tag{49}$$

The exact solution of this equation is

$$a'_n(t) = \exp(-iE_n t/\hbar)a'_n(0) \tag{50}$$

the squared modulus of which is

$$|a'_n(t)|^2 = |a'_n(0)|^2 \tag{51}$$

If the probability of being in the state ϕ_n at time zero is $|a'_n(0)|^2$, then the probability of finding the system in the same state at a later time t is the same if we use the zero potential.

The relationship between the amplitudes a_n and a'_n is given in equation (31). This equation can be rewritten to give

$$a_n(t) = \sum_m \langle \phi_n | \exp(-iq\Lambda/\hbar c)\phi_m \rangle a'_m(t) \tag{52}$$

where $\Lambda = \Lambda(x)$ is an arbitrary differentiable function of space and time. Therefore if $|a'_n(t)|^2$ is constant in time, $|a_n(t)|^2$ in general is not constant in time, unless Λ is a function of the time only. If the system is prepared in the state ϕ_n at time zero, the pseudointeraction in equation (48) causes the system to change to other states at later times. It is only when the potential is chosen to be zero (or the scalar potential is a function of the time only), that the conventional procedure predicts that the state probabilities are constants in time when no field is present. Thus the conventional procedure is gauge dependent. The gauge which gives the correct results when no electromagnetic field is present is the one in which the potential is zero (or the scalar potential is a function of the time only). This choice of potential is called the "preferential gauge" by Leubner and Zoller (1980). The use of the preferential gauge when no electromagnetic field is present is an unwritten rule of quantum mechanics.

On the other hand, the gauge-invariant procedure in Section 6 predicts that in all gauges the probability of finding the system in a given energy eigenstate is a constant if there is no electromagnetic field. The matrix

element in equation (40) for the potential in equation (46) is

$$\langle\psi_n|(-i\hbar c\,\partial_0 + q\,\partial_0\Lambda)\psi_m\rangle = \langle\psi_n'|(-i\hbar c\,\partial_0)\psi_m'\rangle = 0 \tag{53}$$

where ψ_n' is related to ψ_n by equation (41). When the potential is zero as in equation (47), the energy operator in equation (14) reduces to H_0, the unperturbed Hamiltonian in equation (26), and the energy eigenvalue equation in equation (38) reduces to the unperturbed eigenvalue equation in equation (27). Thus the energy eigenvalues $\varepsilon_n = E_n$ and eigenstates $\psi_n' = \phi_n$, which are time independent. The time derivative ∂_0 acting on ψ_m' in equation (53) thus gives zero.

Since the matrix element in equation (40) vanishes, the equation becomes

$$i\hbar\dot{c}_n - E_n c_n = 0 \tag{54}$$

which has the solution

$$c_n(t) = \exp(-iE_n t/\hbar)c_n(0) \tag{55}$$

The squared modulus of equation (55) gives the probability that the system is in the energy eigenstate n (or $[\psi_n]$) at time t,

$$P_n(t) = |c_n(t)|^2 = |c_n(0)|^2 \tag{56}$$

so that $P_n(t)$ is a constant. Equation (43) guarantees that this same result is obtained in all gauges. If the system is prepared in the state n it will stay in the state n for all time if the electromagnetic field is zero. Since the electromagnetic field is classical, this result agrees with our intuition.

8. COMPARISON OF THE GAUGE-INVARIANT AND CONVENTIONAL AMPLITUDES

If the conventional amplitudes are in general gauge dependent, why does the conventional approach often work in practice? This question was answered by Leubner and Zoller (1980) and by Yang (1981). When the *potential* is switched on at some time and off at some subsequent time, the energy operator is equal to the unperturbed Hamiltonian when the potential is zero. The conventional probability amplitude agrees with the gauge-invariant probability amplitude calculated from the same potential when the potential is zero. The sudden switching on or off of the potential can cause a delta function electric field which must be taken into account in the

gauge-invariant procedure. However, if adiabatic switching of the potential is used, there is no delta function contribution to the electric field. For scattering problems the potential is switched on and off adiabatically, and only the probability per unit time that a transition has taken place after an infinite time is required to calculate the cross section. In this case the conventional and the gauge-invariant formulations agree.

The four-potential can be chosen to be

$$a_\mu(x) = \theta(T+t)\theta(T-t)A_\mu(x) \tag{57}$$

for $T>0$, where $A_\mu(x)$ is an everywhere differentiable function. The unit step function $\theta(t)$ is one for positive argument and zero for nonpositive argument. When this potential is used in Section 5, the quantity $|a_n(t)|^2$ does not depend on the time for $t<-T$, $t>T$, when the electromagnetic field and the potential are both zero. The energy operator in equation (14) is

$$\mathcal{E}(a_i) = H_0 \qquad \text{for } t<-T, \qquad t>T \tag{58}$$

so that equation (38) reduces to equation (27). The eigenfunctions $\psi_n(t)$ and eigenvalues $\varepsilon_n(t)$ in equation (38) then become

$$\psi_n(t) = \phi_n, \qquad \varepsilon_n(t) = E_n \qquad \text{for } t<-T, \qquad t>T \tag{59}$$

The amplitude $c_n(t)$ in equation (39) becomes

$$c_n(t) = \langle \phi_n | \psi(t) \rangle \qquad \text{for } t<-T, \qquad t>T \tag{60}$$

which is the same as for a_n in equation (28). Therefore in this case

$$c_n(t) = a_n(t) \qquad \text{for } t<-T, \qquad t>T \tag{61}$$

The "electromagnetic field" $f_{\mu\nu}$ is calculated by using the potential in equation (57) in equation (7), which gives

$$\begin{aligned} f_{\mu\nu} &= \theta(T+t)\theta(T-t)F_{\mu\nu} \\ &\quad + c^{-1}\{\delta(T+t)-\delta(T-t)\}\{\delta_{\mu 0}A_\nu - \delta_{\nu 0}A_\mu\} \end{aligned} \tag{62}$$

The electromagnetic field strength $F_{\mu\nu}$ is defined in equation (7). The "electromagnetic field" $f_{\mu\nu}$ depends on the gauge because a_μ in equation (57) is not a proper potential like A_μ. If a gauge transformation in equation (6) is made on A_μ in equation (57) to give A'_μ, equation (62) would have $\delta_{\mu 0}A'_\nu - \delta_{\nu 0}A'_\mu$ as the coefficient of the delta functions. Thus the potential in

equation (57) does not in general describe a realistic electromagnetic field because of the delta function terms in equation (62).

If the potential A_μ satisfies

$$A_\mu(\mathbf{r}, -T) = A_\mu(\mathbf{r}, T) = 0 \tag{63}$$

for all μ, then the delta function terms in equation (62) vanish. If in addition,

$$F_{\mu\nu}(\mathbf{r}, -T) = F_{\mu\nu}(\mathbf{r}, T) = 0 \tag{64}$$

then the electromagnetic field is continuous. When the potential and the field satisfy equations (63) and (64), respectively, then the equality in equation (61) holds for a realistic situation in which the electromagnetic field $F_{\mu\nu}$ is switched on at time $-T+\varepsilon$ and switched off at time $T-\varepsilon$, where $\varepsilon \to 0$.

When adiabatic switching is used, the potential and electromagnetic field are assumed to satisfy equations (63) and (64). In addition the potential and field are assumed to be switched on "slowly" after time $-T$ and switched off "slowly" before time T. The limit as T approaches infinity can be taken at the appropriate place in the calculation of cross sections. The cross sections obtained by the conventional procedure and the gauge-invariant procedure are the same because of equation (61). The equality is valid in all orders of perturbation theory.

9. CONCLUSION

The gauge invariance of the Dirac equation for an electron in an external classical electromagnetic field is reviewed in this paper. In order to ensure manifest gauge invariance, gauge-invariant operators are defined. The energy operator, not the unperturbed Hamiltonian, is used to obtain manifestly gauge-invariant probability amplitudes. The conventional procedure, based on the unperturbed Hamiltonian, gives gauge-dependent probability amplitudes in general. Both the gauge-invariant and the conventional approaches are applied to the special case of a Dirac particle in a zero electromagnetic field. The conventional approach predicts that if the potential is the four-gradient of an arbitrary scalar function, the probability that the system is in a given state changes in time, even though the electromagnetic field is zero. On the other hand, the gauge-invariant formulation predicts that in all gauges the probability that the system is in a given state is a constant, which agrees with our intuition.

In many problems the adiabatic switching hypothesis is made, so that before the potential (and the field) is switched on and after the potential (and the field) is switched off, the gauge-invariant and conventional probability amplitudes agree with each other (Leubner and Zoller, 1980). The adiabatic switching hypothesis is most often used in scattering problems so that cross sections calculated by the two approaches agree in all orders of perturbation theory. However, there are other effects, like the AC Stark shift, in which the field remains on while measurements are made. In the nonrelativistic case the conventional and gauge-invariant formulations give different AC Stark shifts (Kobe, 1982), and similar results would be obtained if the Dirac equation were used. In general, eigenstates of the energy operator are required to calculate probability amplitudes. When the energy operator is different from the unperturbed Hamiltonian, the gauge-invariant procedure must be used to calculate probability amplitudes.

In the special case that the electric dipole approximation can be made for the hydrogen atom (i.e., the wavelength is long compared to the atomic dimensions and magnetic effects are negligible), the potentials can be chosen (Kobe, 1978) such that $\mathbf{A}' = 0$, $A_0' = -\mathbf{E}(0,t)\cdot\mathbf{r}$, where $\mathbf{E}(0,t)$ is the electric field at the origin (at the atom). In this case the kinetic momentum reduces to the canonical momentum, and the energy operator reduces to the unperturbed Hamiltonian. The matrix element of the interaction operator in Equation (40) becomes $\langle\psi_n'|c\pi_0'\psi_m'\rangle = \langle\phi_n|\{-q\mathbf{E}(0,t)\cdot\mathbf{r}\}\phi_m\rangle$. If the unperturbed Hamiltonian H_0 is used then $-q\mathbf{E}(0,t)\cdot\mathbf{r}$ should be used as the interaction operator to obtain probability amplitudes (Yang, 1976; Kobe and Smirl, 1978). However, if the adiabatic switching hypothesis is used, the conventional and the gauge-invariant amplitudes are equal when the potential (and field) is zero. Dirac (1958) gave a direct proof that in second-order perturbation theory for the quantized electromagnetic field in the electric dipole approximation the Kramers–Heisenberg formula calculated using $-q\mathbf{E}(0,t)\cdot\mathbf{r}$ and the one calculated using $\mathbf{A}(0,t)\cdot\mathbf{p}$ plus A^2 (constants omitted) are equal.

In general the electromagnetic field can remain on during a measurement, so the energy operator and the unperturbed Hamiltonian differ. In this case, the gauge-invariant formulation should be used to ensure gauge-invariant results.

ACKNOWLEDGMENTS

I would like to thank Dr. K.-H. Yang for many helpful discussions on gauge invariance, for reading the manuscript and making helpful suggestions, and for his encouragement. This work was supported in part by a grant from the North Texas State University Faculty Research Fund.

REFERENCES

Bjorken, J. D., and Drell, S. (1964). *Relativistic Quantum Mechanics*, Chap. 1. McGraw-Hill, New York.

Bohm, D. (1951). *Quantum Theory*, pp. 357–358, Prentice-Hall, New York.

Cohen-Tannoudji, C., Diu, B., and Laloë, F. (1977). *Quantum Mechanics*, pp. 315–328. John Wiley and Sons, New York.

Dirac, P. A. M. (1928a). *Proceedings of the Royal Society of London*, **A117**, 610.

Dirac, P. A. M. (1928b). *Proceedings of the Royal Society of London*, **A118**, 351.

Dirac, P. A. M. (1950). *Nuovo Cimento (Series 9)*, **7**, 925.

Dirac, P. A. M. (1951). *Proceedings of the Royal Society of London*, **A209**, 291.

Dirac, P. A. M. (1952). *Annales de l'Institut Henri Poincaré*, **13**, 1.

Dirac, P. A. M. (1955). *Canadian Journal of Physics*, **33**, 650.

Dirac, P. A. M. (1958). *The Principles of Quantum Mechanics*, 4th ed., pp. 244–248, pp. 34–37, Oxford University Press, London.

Dirac, P. A. M. (1965). *Physical Review*, **139**, B684.

Dirac, P. A. M. (1970). *Physics Today*, April, **23**, 29.

Epstein, S. T. (1979). *Chemical Physics Letters*, **65**, 417.

Feynman, R. P. (1962). *Quantum Electrodynamics*, pp. 34–55. W. A. Benjamin, Inc., New York.

Kobe, D. H. (1978). *International Journal of Quantum Chemistry Symposium*, **12**, 73.

Kobe, D. H. (1983). (to be published).

Kobe, D. H., and Smirl, A. L. (1978). *American Journal of Physics*, **46**, 624.

Kobe, D. H., and Yang, K.-H. (1980). *Journal of Physics A: Mathematical and General*, **13**, 3171.

Leubner, C., and Zoller, P. (1980). *Journal of Physics B: Atomic and Molecular*, **13**, 3613.

Rose, M. E., and Welton, T. A. (1952). *Physical Review*, **86**, 432.

Rose, M. E. (1961). *Relativistic Electron Theory*, pp. 117–119. John Wiley and Sons, New York.

Sakurai, J. J. (1967). *Advanced Quantum Mechanics*, Chap. 3. Addison-Wesley, Reading, Massachusetts.

Schectman, R. M., and Good, R. H., Jr. (1956). *American Journal of Physics*, **25**, 219.

Schwinger, J. (1951). *Physical Review*, **82**, 664.

Yang, K.-H. (1976). *Annals of Physics*, **101**, 62.

Yang, K.-H. (1981). *Physics Letters*, **81A**, 125.

Yang, K.-H. (1982). *Journal of Physics A: Mathematical and General*, **15**, 437.