International Journal of Theoretical Physics, Vol. 19, No. 7, 1980

Quantum Interference and the Classical Limit

J. Anandan

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260 and Institute for Theoretical Physics, [1] *University of Maryland, College Park, Maryland 20742*

Received May 16, 1978

A relationship between the quantum phase shift due to an external field in an (infinitesimal) two-beam interference experiment and the classical transverse acceleration, previously given by the author, is generalized covariantly and the phase shift associated with the longitudinal acceleration is deduced. This relation implies that the mass of a spinless elementary particle is a constant of motion, even in the presence of external fields in a curved space-time. Also according to this relation, the infinitesimal Aharonov–Bohm effect is equivalent to the Lorentz force law. The case of a charged particle moving in the electromagnetic field due to electric and magnetic charges obeying Dirac's quantization condition is treated and a prescription is given for constructing its quantum mechanical wave function directly from the classical equation of motion, using path integrals, without requiring a knowledge of the Lagrangian or Hamiltonian of the interaction. Generalization to an arbitrary gauge field is briefly considered and a theorem is proven which implies that the holonomy transformations (path-ordered operators that parallel transport around closed curves) at an arbitrary space-time point contain all the gauge-invariant information of a gravitational or gauge field connection.

1. INTRODUCTION

The importance of the phase difference between two interfering coherent beams in determining the motion of spinless particles has been recognized implicitly and explicitly by many authors. Indeed, this forms the basic physical idea behind Feynman's path integral formulation of quantum mechanics. According to Feynman (1948) the amplitude $K(x,x')$ for a particle to propagate from one space-time point x' to another x can

[1]Present address. Address after September 1980: Department of Mathematics, University of California, Berkeley, California 94720.

0020-7748/80/0700-0537$03.00/0

be expressed as a sum of the amplitudes $\exp[(i/\hbar)S_\gamma(x,x')]$ associated with all possible paths γ connecting x' to x:

$$K(x,x')\sim\sum_\gamma \exp\left[(i/\hbar)S_\gamma(x,x')\right] \tag{1.1}$$

where $S_\gamma/\hbar$ may be thought of as the phase associated with the path γ.

S_γ, according to Feynman, is the classical action associated with the path γ. This prescription therefore provides a method of quantizing a classical theory that can be obtained from an action principle. Conversely, if the propagator $K(x,x')$ for the quantum theory is given in the form of (1.1), in the classical limit when $S_\gamma/\hbar$ becomes very large, as pointed out by Feynman, the path for which $S\gamma$ is stationary and its neighboring paths make the major contribution to the amplitude. This path may therefore be regarded as the classical trajectory. Thus Feynman's formalism provides a clear intuitive connection between the classical theory and the quantum theory in cases where the classical theory is obtainable from an action principle.

In a previous paper (Anandan, 1977) a relationship between the phase shift for interference between beams along two paths enclosing an infinitesimal spatial area and the classical transverse acceleration, due to an external field, was given. This raises the question of whether Feynman's scheme of quantization can be extended to more general cases where a system has a classical equation of motion which may or may not be derivable from an action principle. Apart from raising this possibility, the above-mentioned relation has some intrinsic merit because it provides a very intuitive connection between classical physics and quantum physics and also treats in a unified manner the phase shifts due to different interactions. It therefore deserves further investigation, which is the major objective of the present paper.

An outline of the present paper is as follows. In Section 2, the above-mentioned relationship between the phase shift in quantum interference and the classical transverse acceleration is covariantly generalized in curved space-time and the phase shift associated with the longitudinal acceleration is deduced. This general relationship implies that the mass of a spinless elementary particle is a constant of motion even in the presence of external fields. It is shown also that a particle which undergoes the Aharonov–Bohm (1959) effect[2] in quantum interference must obey the

[2]By the Aharonov–Bohm effect in this paper, we simply mean the phase shift in quantum interference due to the electromagnetic field, which need not vanish along the interfering beams. This is a generalization of the usual Aharonov–Bohm effect in which the field is assumed to vanish along the beams.

Lorentz force law in the classical limit and vice versa (infinitesimally), in an external field described by a skew symmetric tensor F_{ab} that need *not* obey Maxwell's equations. In particular this allows, in Section 3, for the formulation of the quantum theory of a charged particle, in the presence of an electromagnetic field due to magnetic and electric charges obeying Dirac's quantization rule, directly from the classical equation of motion without requiring a classical Lagrangian or action principle for the interaction. In the process of doing this we also give a derivation of Dirac's quantization rule that does not introduce strings. A possible generalization of our propagator to arbitrary gauge fields using path-ordered operators around closed curves (holonomy transformations) is suggested. A theorem is proved in the Appendix which implies that these path-ordered operators contain all the gauge-invariant information of the field.

2. THE CORRESPONDENCE PRINCIPLE

We shall call a system which exhibits both wave and particle properties a quantum mechanical system, as opposed to systems that are purely wave or purely particle, which are examples of classical systems. An electron, for instance, is a quantum mechanical system: when it passes through a cloud chamber it leaves evidence of a particle trajectory, whereas on the other hand, it undergoes interference and diffraction which are characteristic of waves. The main purpose of this paper is to relate these two aspects of a quantum mechanical system by investigating the connection between interference and classical trajectories. We shall begin by assuming that space-time is a four-dimensional differentiable manifold.[3] In the absence of any external fields (other than gravity) we shall also suppose that the motion of a quantum mechanical system is governed by a propagator of the form (1.1), which may be regarded as a formal way of stating Huygen's principle, which determines the evolution of the associated wave by its interference with itself. The difference between the phases associated with two arbitrary paths joining x and x' in (1.1) will be called the phase difference between these two paths. We shall assume that

[3]It may appear that it is not necessary to make this assumption. We could instead begin with space-time M as a set of points called events. Experience suggests that M has a topological structure such that the wave function is a continuous function on M with respect to this topology, for all states of motion of any particle. The four dimensionality of space-time is a consequence of this topology. It also appears from experience that M has a differentiable structure such that the wave functions are differentiable functions on M. This makes M a differentiable manifold. But this procedure is unsatisfactory because the measurement process shows that the value of the wave function at an event has no reality. See Anandan (1980).

an external field changes the phase difference $\Delta\chi$ in the absence of any field to $\Delta\chi+\Delta\phi$, where $\Delta\phi$ is called the phase shift due to the field. This is a weaker assumption than the usual assumption which provides a prescription for the change in phase along each path joining x and x' due to the external field. As we shall see in the next section, the former assumption is sufficient to determine the modifications of the motion of the system in the presence of the external field. Moreover, for the electromagnetic field, for instance, it is gauge invariant, unlike the latter assumption. This will enable us to obtain the new quantum mechanical motion directly from the classical equation of motion, which is also gauge invariant. In order to do this we shall first obtain a relation between the phase shift for two paths infinitesimally close to each other and the classical equation of motion.

We define a classical trajectory to be a curve such that $\Delta\chi+\Delta\phi$ between this curve and all neighboring curves is zero to first order in the variation. In the absence of any external field other than gravity ($\Delta\phi=0$), the classical trajectories associated with different massive particles are the same, which is one way of stating the equivalence principle in the classical limit. There exists also another class of classical trajectories associated with massless particles, which is also independent of which massless particle is being considered. This suggests that these classical particles represent an objective geometrical structure of space-time. It is also known that under reasonable assumptions there exists a pseudo-Riemannian metric g_{ab} of signature $(+---)$, unique up to an overall factor, such that the above two classes of curves are, respectively, the timelike and null geodesics with respect to this metric. (Weyl, 1921; Marzke and Wheeler, 1964; Ehlers, Pirani, and Schild, 1972). The existence of such a metric will be assumed from now on.

Now, an external field would give a nonzero phase shift $\Delta\phi$ and hence cause a deviation of the classical trajectory from the geodesic curve. For simplicity consider first the classical path of a particle in three-dimensional space, as seen from a local inertial frame. This path may be regarded as an appropriate projection on a spacelike hypersurface of its classical trajectory in space-time. Hence in the absence of an external field this path will be locally a straight line. Suppose that an external field is now "turned on," causing a curvature in the classical path of the particle. Then according to the above considerations this curvature is associated with the phase shift in quantum interference due to the external field. Using essentially this idea, the following relativistic relation between the phase shift and the classical transverse acceleration was obtained in a previous paper (Anandan, 1977; see also Feynman, 1964):

$$\Delta\phi_{\perp}=\frac{mA\gamma(dv_{\perp}/dt)}{\hbar v} \tag{2.1}$$

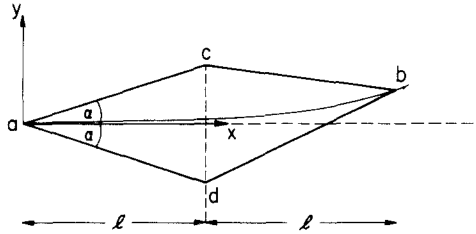

Fig. 1. The connection between the phase shift in a two-beam interference and the classical transverse acceleration due to an external field. The external field causes a phase shift between the interfering beams shown schematically by *acb* and *adb* so that constructive interference occurs at *b* which is on the classical path *ab* through *a*. The spatial axes *ax*, *ay* emphasize that the classical path is the projection on the x–y plane of a classical trajectory in space-time.

where $\Delta\phi_\perp$ is the phase shift due to the external field, associated with two curves that span the infinitesimal area A, $dv_\perp/dt$ is the component of the acceleration perpendicular to the velocity, in the plane of interference which contains **v**, m is the mass, and $\gamma=(1-v^2/c^2)^{-1/2}$ (see Figure 1). By A being infinitesimal, we mean that (2.1) is valid only in a limiting sense. A more precise statement of (2.1) is

$$\frac{mA\gamma(dv_\perp/dt)}{v\Delta\phi_\perp}\to\hbar \text{ as } l\to 0 \text{ and } \alpha\to 0 \tag{2.1$'$}$$

where l is the length and α the angle indicated in Figure 1. All equations in which $\Delta\phi$ appears in this section should be understood as being valid in such a limiting sense. There can exist, however, special cases in which these equations are valid for a finite area so long as the variation of the field over this area is negligible. An example of this kind is the Aharonov–Bohm effect, which will be considered later.

There is also a phase shift $\Delta\phi_\parallel$ corresponding to the longitudinal acceleration $dv_\parallel/dt$ (component of acceleration parallel to velocity), for an appropriate experiment done in space-time (Figure 2). The covariant generalization of (2.1) in curved space-time, which will give both phase shifts and include also the case when the particle is massless, as will be seen later, is

$$\Delta\phi v^b=\frac{c}{\hbar}\,d\sigma^{ab}\frac{Dp_a}{Ds} \tag{2.2}$$

where $v^a=dx^a/ds$ is the 4-velocity, p^a is the 4-momentum, D/Ds denotes covariant differentiation with respect to an affine parameter along the worldline of a classical particle, and $\Delta\phi$ is the phase shift due to the

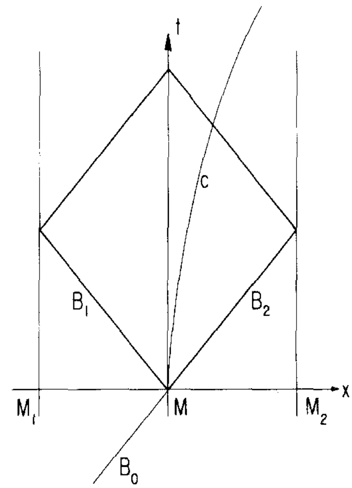

Fig. 2. A schematic diagram of an experiment in which eam B_0 is split into two beams B_1 and B_2 by a half-reflecting mirror whose world line is M. B_1 and B_2 after reflection by the mirrors M_1 and M_2 interfere on M. M_1, M_2, and M are at rest in an inertial coordinate system whose time axis and a distance axis are denoted by t, x. An external field is present so as to cause a phase shift $\Delta\phi$ between the interfering beams and a corresponding longitudinal acceleration in the classical trajectory C.

external field between the interfering beams spanning the infinitesimal surface element represented by $d\sigma^{ab}$, which must contain v^a (Figure 3). Since $d\sigma^{ab}$ contains v^a, we can write

$$d\sigma^{ab} = u^{[a}v^{b]} \tag{2.3}$$

where u^a is an infinitesimal vector whose direction is arbitrary so long as it is not parallel to v^a. In special cases, $\Delta\phi$ in (2.2) can be identified with the phase shift in a suitable two-beam interference experiment. Equation (2.2) together with (2.3) can be regarded as a *correspondence principle* between classical physics and quantum physics.

It follows from (2.2) and (2.3) that

$$v^a \frac{Dp_a}{Ds} = 0 \tag{2.4}$$

Hence the 4-force $f^a = Dp^a/Ds$ experienced by a particle which obeys (2.2) must always be normal to the 4-velocity. For spinless elementary particles we may also suppose that $p^a = mv^a$, where m by definition is the mass of

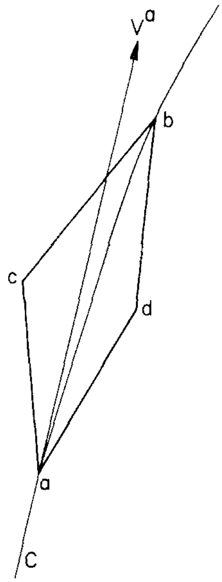

Fig. 3. C represents the classical trajectory of a particle and v^a is the 4-velocity at a point a on C. The curvature of the classical trajectory is related by equation (2.2) to the phase shift between interfering beams acb, adb, where b is a point on the classical trajectory and a, b, c and d are infinitesimally close to each other.

the particle. Then from (2.4),

$$Dm/Ds = 0 \tag{2.5}$$

i.e., the mass of a spinless elementary particle should be a constant of motion.[4]

If t^a is any vector such that $v^a t_a \neq 0$, then from (3.2),

$$\Delta\phi = \frac{c}{\hbar}\,\frac{d\sigma^{ab}(Dp_a/Ds)t_b}{v^c t_c} \qquad (v^c t_c \neq 0) \tag{2.6}$$

If the particle has a time like 4-velocity v^a, we can choose t^a to be v^a and obtain

$$\Delta\phi = \frac{c}{\hbar}\,d\sigma^{ab}\frac{Dp_a}{Ds}v_b \qquad (v^c v_c = 1) \tag{2.7}$$

[4]In flat space-time, in the absence of any external field, Poincaré invariance guarantees that mass is a constant of motion. This need not be the case if there is a gravitational or some other field. (2.2) implies that for a spinless elementary particle (i.e., a particle whose motion is described by a single complex scalar function of space-time), the mass must be a constant. It is possible to generalize (2.2) so that the mass is not a constant of motion. But it would be artificial to do so.

If tachyons exist it is possible in principle to do an interference experiment with them. In this case,

$$\Delta\phi = -\frac{c}{\hbar} d\sigma^{ab} \frac{Dp_a}{Ds} v_b \qquad (v^c v_c = -1) \tag{2.8}$$

For massless particles, however, $\Delta\phi$ must be obtained from the more general relation (2.2) or (2.6).

From (2.2) we shall derive now (2.1) and also the phase shift due to longitudinal acceleration mentioned earlier. Consider first a massive spinless elementary particle for which (2.7) is valid and choose a local inertial coordinate system in which D/Ds is the same as the ordinary derivative d/ds. Also this coordinate system can be chosen such that, without loss of generality, in this coordinate system

$$v^a = \gamma\left(1, \frac{v}{c}, 0, 0\right)$$

$$\frac{dp^a}{ds} = m\frac{dv^a}{ds} = \frac{m\gamma^2}{c^2}\left(\gamma^2 \frac{v}{c}\frac{dv_{\parallel}}{dt}, \gamma^2 \frac{dv_{\parallel}}{dt}, \frac{dv_{\perp}}{dt}, 0\right) \tag{2.9}$$

where $\gamma = (1 - v^2/c^2)^{-1/2}$ and $dv_{\parallel}/dt$ is the component of acceleration parallel to velocity. Substitute (2.9) into (2.7). Since $d\sigma^{ab}$ contains v^a, $d\sigma^{21}/d\sigma^{20} = v/c$, and so one obtains

$$\Delta\phi = \frac{mA\gamma}{\hbar v}\frac{dv_{\perp}}{dt} + \frac{m\gamma^3}{\hbar c}\Sigma\frac{dv_{\parallel}}{dt}$$

where $A = d\sigma^{12}$ and $\Sigma = d\sigma^{01}$. Noticing now that $m\gamma(dv_{\perp}/dt)$ and $m\gamma^3(dv_{\parallel}/dt)$ are the transverse and longitudinal components of $d\mathbf{p}/dt$, we can write

$$\Delta\phi = \Delta\phi_{\perp} + \Delta\phi_{\parallel} \tag{2.10}$$

where

$$\Delta\phi_{\perp} = \frac{A(dp_{\perp}/dt)}{\hbar v} \tag{2.11}$$

and

$$\Delta\phi_{\parallel} = \frac{\Sigma(dp_{\parallel}/dt)}{\hbar c} \tag{2.12}$$

It can be easily verified that (2.10)–(2.12) follow also from (2.8) and from the more general relation (2.6) so that they are valid for massless particles and tachyons as well. When $p^a = mv^a$, m being a nonzero constant, (2.11) is the same as (2.1). Thus we have derived (2.1) as well as the new relation (2.12) which gives the phase shift due to longitudinal acceleration (Figure 2), from (2.2). Thus (2.2) is a covariant generalization of (2.1).

As an example consider the electromagnetic field F_{ab} for which the phase shift is

$$\Delta\phi = \frac{e}{2\hbar} F_{ab} d\sigma^{ab} \tag{2.13}$$

which is the covariant formulation of the Aharonov–Bohm effect for an infinitesimal interference loop. It follows from (2.13), (2.2), and (2.3), on remembering (2.4) and using the arbitrariness of the direction of u^a, that

$$\frac{Dp^a}{Ds} = \frac{e}{c} F^a_b v^b \tag{2.14}$$

which is the Lorentz force law. Conversely, from (2.14), on using (2.2) and (2.3), we can deduce (2.13).

3. AHARONOV–BOHM EFFECT, MAGNETIC MONOPOLES, AND PATH INTEGRALS

In the end of the last section we showed that according to the correspondence principle (2.2), the infinitesimal Aharonov–Bohm effect (2.13) is equivalent to the Lorentz force law (2.14). The six components of the electromagnetic field F_{ab} that are known to exist are, in this view, to be associated with the six components of an arbitrary surface element $d\sigma^{ab}$. Indeed the requirement that the phase shift $\Delta\phi$ be a scalar, dependent on $d\sigma^{ab}$, implies that the simplest local interaction a particle can have is with a skew symmetric rank-2 tensor field F_{ab}. We shall continue to refer to F_{ab} as the electromagnetic field even though it does not have to satisfy the usual Maxwell's equations corresponding to the absence of magnetic monopoles. But we shall show in this section that it is easy to give a prescription for constructing the wave function for a spinless particle undergoing this interaction, provided F_{ab} satisfies a condition which is equivalent to Dirac's quantization rule for magnetic monopoles.

$d\sigma^{ab}$ in (2.13) is infinitesimal, whereas in reality, the area of a surface spanned by the interference loop will be finite. Given such a surface S, divide it into tiny areas S_α (Figure 4). Let $\Delta\phi_\alpha$ denote the phase shift for quantum interference around S_α and define $\Delta\phi_s \equiv \Sigma_\alpha \Delta\phi_\alpha$. In the present case, using our correspondence principle and the Lorentz force law, $\Delta\phi_\alpha$ is given by (2.13) for an infinitesimal area S_α, as shown earlier. So, as the

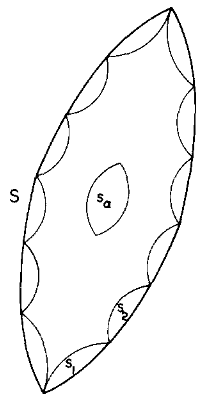

Fig. 4. A surface S spanned by an interference loop can be divided into infinitesimal surface elements S_α. The relationship between the phase shift for an interference experiment around S and the phase shifts for those around S_α is given by assumption A.

number of areas S_α tend to infinity while the dimensions of the areas tend to zero, we obtain

$$\Delta\phi_s = \frac{e}{2\hbar}\int_S F_{ab}\,d\sigma^{ab} \tag{3.1}$$

A reasonable assumption that may give the phase shift $\Delta\phi$ in interference is $\Delta\phi=\Delta\phi_s$. But the position of interference fringes are determined by $e^{i\Delta\phi}$. So a more general, experimentally verifiable assumption, is the following. Assumption A: Let $\Delta\phi$ be the phase shift due to an external field for quantum interference around a loop γ. If S is any surface spanned by γ, then $e^{i\Delta\phi}=e^{i\Delta\phi_s}$, where $\Delta\phi_s$ was defined in the paragraph preceding (3.1).

Assumption A together with (3.1) is the Aharonov–Bohm effect. The experimental confirmation of this effect (Chambers, 1960), can therefore be regarded as evidence for assumption A and the correspondence principle which gave (3.1) from the Lorentz force law. Now let S' be another surface spanned by γ. Then by assumption A, $e^{i\Delta\phi_s}=e^{i\Delta\phi_s'}$, or equivalently

$$\Delta\phi_s-\Delta\phi_{s'}=\frac{e}{2\hbar}\int_{S\cup S'} F_{ab}\,d\sigma^{ab}=2\pi n, \qquad n=\text{integer} \tag{3.2}$$

where $S\cup S'$ represents the closed surface formed from S and S'. By generalized Stoke's theorem

$$\frac{1}{2}\int_{S\cup S'} F_{ab}\,d\sigma^{ab}=\frac{1}{2}\int_V \varepsilon^{abcd}\nabla_d F_{bc}\,dV_a=\int_V 4\pi\,{}^*j^a\,dV_a$$

where V is the volume enclosed by $S \cup S'$ and $^*j^a$ is the "magnetic current" defined by the generalized Maxwell's equations:

$$\frac{1}{2}\varepsilon^{abcd}\nabla_d F_{bc} = 4\pi^* j^a \tag{3.3}$$

On defining the "magnetic charge" $g = \int_V {}^*j^a dV_a$ (3.2) therefore reads

$$eg = \frac{n\hbar}{2}, \qquad n = 0, \pm 1, \ldots \tag{3.4}$$

which is Dirac's quantization rule (Dirac, 1931, 1948).

To understand the physical meaning and implications of assumption A, it is necessary to consider the meaning of $\Delta\phi$. It is sometimes stated that only $e^{i\Delta\phi}$ and not $\Delta\phi$ can be measured experimentally. This statement, however, needs to be qualified, as we shall see now. Suppose an interference experiment is done, first in the absence of an external field, which is then gradually "turned on." The phase shift $\Delta\phi$, which in the absence of the field may be made zero by definition, will gradually increase in magnitude. By observing the shift in the interference fringes, since the process is a continuous one, $\Delta\phi$ can be measured. Thus even though at any given instant only $e^{i\Delta\phi}$ can be observed, the knowledge of the history of the system provides a definite prescription to obtain $\Delta\phi$ at least in special cases.

As a specific example, consider an interference experiment with electrons in which the interference loop γ is in the $X-Y$ plane of a Cartesian coordinate system $0\ X\ Y\ Z$. S and S' are two surfaces spanned by γ such that the origin 0 is inside the closed surface formed from S and S'. Suppose that γ is far removed from other charges (or is shielded by a conductor) so that the electromagnetic field in its vicinity is zero. The phase shift $\Delta\phi$ is then zero by definition. Now suppose that a particle with magnetic charge g with its initial position very far away from 0 on the positive z axis is gradually brought towards 0 along the z axis. Alternatively, the electromagnetic field can be literally turned on by creating a monopole antimonopole pair and bringing the monopole towards 0. Now S and S' can be divided into infinitesimal surface elements S_α and S_β'; let $\Delta\phi_\alpha, \Delta\phi_\beta'$ denote the phase shifts in interference experiments around these surface elements, due to bringing g towards 0. $\Delta\phi, \Delta\phi_\alpha, \Delta\phi_\beta'$ are experimentally measurable as described in the previous paragraph. Hence $\Delta\phi, \Delta\phi_s \equiv \Sigma_\alpha \Delta\phi_\alpha$ and $\Delta\phi_{s'} \equiv \Sigma_\beta \Delta\phi_\beta'$ are physically meaningful quantities since they can be determined by a well-defined operational procedure.

The electromagnetic field F_{ab} can be determined operationally using (2.13). This can be done more generally even when the history of the charges producing the field is unknown, by assuming that the magnitude of phase shift is less than π for an infinitesimal loop and thereby eliminate the ambiguity associated with adding an integer multiple of 2π. Alternatively F^{ab} can be determined using (2.14) in the classical limit. Once F^{ab} is determined, notice that the magnetic charge considered here, *by definition*, satisfies (3.3). We shall *not*, however, assume now that g satisfies (3.4). It follows from (3.1) and (3.3) that when g is outside the closed surface, $\Delta\phi_s = \Delta\phi_{s'}$. It is reasonable now to assume that $\Delta\phi = \Delta\phi_s = \Delta\phi_{s'}$ which is experimentally verifiable if magnetic monopoles exist. Before g reaches the origin, it crosses one of the two surfaces, say, S'. Suppose $S_\alpha{}'$ for some α is the surface element that g crosses. Then there will be sudden change in $\Delta\phi'_\alpha$ according to (2.13), and consequently such a change will also occur in $\Delta\phi_{s'}$. But no such sudden change will occur in $\Delta\phi$ or $\Delta\phi_s$. So when g is inside the closed surface we should have $\Delta\phi = \Delta\phi_s \neq \Delta\phi_{s'}$. When g is at the origin, $\Delta\phi = \Delta\phi_s = 2\pi eg/\hbar$, $\Delta\phi_{s'} = -2\pi eg/\hbar$ because of (3.3) and (3.1). If g continues its journey and crosses S as well, then $\Delta\phi \neq \Delta\phi_s$, $\Delta\phi \neq \Delta\phi_{s'}$, in general. But if g is pulled back to 0 again along the Z axis, then $\Delta\phi = \Delta\phi_s = 2\pi eg/\hbar$ again. Suppose that the experiment is repeated, this time with g initially at a very large distance away from 0 along the negative Z axis. If g is brought to 0 along the Z axis, by a similar argument, $\Delta\phi = \Delta\phi_{s'} = -2\pi eg/\hbar$, $\Delta\phi_s = 2\pi eg/\hbar$. The difference in the values of $\Delta\phi$ for the two ways in which g is brought to 0 is one way of realizing the fact that magnetic monopoles violate reflection symmetry. If we require reflection symmetry of course, magnetic monopoles cannot exist (g in above experiment should be zero). But on the other hand if magnetic monopoles exist, thus violating reflection symmetry, it is not *a priori* necessary that the position of the interference fringes for the two cases considered above should be the same. Although in both cases g is at 0 during some time interval when the interference fringes are being observed, the result of the experiment could depend on the history of g.[5] If the possibility of some nonlocal interaction that will make the positions of the interference fringes dependent on the history of the charges producing the field, is excluded, then we must have $e^{i\Delta\phi_s} = e^{i\Delta\phi_{s'}}$, which leads to Dirac's quantization via (3.2). This provides a physical meaning to assumption A.

If magnetic monopoles do not exist, then always $\Delta\phi = \Delta\phi_s = \Delta\phi_{s'}$ and the distinction between these quantities is unnecessary. Assumption A is the simplest generalization of this situation to include magnetic monopoles,

[5]This implies that the quantum mechanical motion of a test particle in the field of g can depend on the history of g. But it is interesting that the equation of motion in the classical limit will be independent of the history of g, according to our assumptions. This can be seen from (2.2) and the fact that the infinitesimal phase shift $\Delta\phi$ in this equation will be independent of the history of g ($\Delta\phi \to 0$ as $d\sigma^{ab} \to 0$).

which must then satisfy Dirac's condition (3.4). Physically this generalization represents the requirement that the predicted shift in the interference pattern due to the Aharonov–Bohm effect must be independent of which surface S is being considered, even in the presence of magnetic monopoles. We have shown that further generalizations that permit violations of (3.4) are possible, but they require more complicated nonlocal interactions between magnetic and electric charges.

The derivation of (3.4) given here is analogous to that of Dirac (1931), but is more general because it was done in curved space-time. The usual derivation of the Aharonov–Bohm effect assumes the existence of a vector potential A_a such that $F_{ab}=\nabla_{[b}A_{a]}$. But if this is true in any neighborhood then Maxwell's equations $\epsilon^{abcd}\nabla_d F_{bc}=0$ must be valid in that neighborhood. Dirac showed that by introducing line singularities in A_a, the more general equation (3.3) will hold provided the subsidiary condition (3.4) is imposed to ensure consistency with quantum mechanics. In our derivation of the Aharonov–Bohm effect, however, the vector potential A_a was not needed. Instead we used the correspondence principle (2.2), which gave (2.13) directly in terms of the observable field F_{ab} from the classical equation of motion. We needed the skew symmetry of F^{ab} but did not require that F^{ab} satisfy Maxwell's equations. This enabled our derivation of the infinitesimal Aharonov–Bohm effect, to permit the existence of magnetic currents. The further assumption A then yielded the experimentally observed Aharonov–Bohm effect and the Dirac's quantization rule (3.4). The latter, therefore has also been obtained entirely in terms of the observable F_{ab} and avoids altogether the use of the vector potential A_a. Hence we do not have the problems associated with the singularities of A_a which confronted Dirac.[6] In this respect, this derivation is similar to that of Cabibbo and Ferrari (1962). But it has the advantage that it will now lead to a well-defined wave function, unlike the path-dependent wave function used by these authors, by means of a procedure very similar to that of Feynman (1948).

Let $\{\gamma_r\}$ be the set of paths joining two space-time points x' and x, and S_r a surface spanned by a fixed path γ_1 and γ_r (r includes 1) in a simply connected region. The Aharonov–Bohm effect and Feynman's "democracy of all paths" suggest that the amplitude for a particle with electric charge e to go from x' to x is

$$\tilde{K}(x,x')\sim\sum_r \exp\left(\frac{ie}{2\hbar}\int_{S_r}F_{ab}\,d\sigma^{ab}\right)\kappa_r \tag{3.5}$$

[6]A method of avoiding singularities in A_a has been given by Wu and Yang (1975) in the integral formalism of electromagnetism.

where κ_r are the amplitudes in the absence of the field F_{ab}. (3.4) is assumed valid so that S_r can be any surface spanned by γ_1 and γ_r not containing a point at which F_{ab} is singular. $\tilde{K}(x,x')$ depends on the choice of γ_1 which can be determined for arbitrary x,x' as follows. Define a field of "reference paths" $\gamma(x)$ joining a fixed "reference point" x_0 to x such that $\gamma(x)$ varies continuously with x, e.g., $\gamma(x)$ in flat space-time can be the straight line joining x_0 to x; in curved space-time it can be, at least locally, a geodesic joining x_0 to x. Given two points x, x', choose γ_1 in (3.5) to be the curve formed by combining $\gamma(x)$ and $\gamma(x')$. If $\tilde{\psi}(x')$ is the amplitude (a continuous function) on a spacelike surface σ then the amplitude at a subsequent point x is postulated to be

$$\tilde{\psi}(x)=\int_\sigma \tilde{K}(x,x')\tilde{\psi}(x')\,dx' \tag{3.6}$$

This formalism has the advantage of using only the observable F_{ab}. But it can be easily shown to be equivalent to Feynman's formalism in the absence of magnetic monopoles. In this case F_{ab} can be written as the curl of a singularity free vector potential A_a and Feynman's propagator (1.1) is

$$K(x,x')\sim\sum_r \exp\left(i\frac{e}{\hbar}\int_{\gamma_r} A_a\,dx^a\right)\kappa_r \tag{3.7}$$

Now define

$$\Lambda(x)\equiv\int_{\substack{x_0\\ \gamma(x)}}^{x} A_a\,dx^a \tag{3.8}$$

Then from (3.5), (3.7), and (3.8),

$$K(x,x')=\exp\left(i\frac{e}{\hbar}\Lambda(x)\right)\tilde{K}(x,x')\exp\left(-i\frac{e}{\hbar}\Lambda(x')\right) \tag{3.9}$$

Therefore if

$$\psi(x)\equiv\exp\left(i\frac{e}{\hbar}\Lambda(x)\right)\tilde{\psi}(x) \tag{3.10}$$

then the equation $\psi(x)=\int_\sigma K(x,x')\psi(x')\,dx'$ in Feynman's formalism is equivalent to (3.6). If $F_{ab}=0$ everywhere,[7] then from (3.8) $\nabla_a\Lambda=A_a$. Thus in this case (3.10) is just the gauge transformation that makes the new

[7]More generally, if $F_{ab}=0$ in a neighborhood V such that for every $x\in V$, V contains $\gamma(x)$, then $\nabla_a\Lambda=A_a$ in V. Then (3.10) is the gauge transformation that makes the new potential $A_a-\nabla_a\Lambda$ vanish in V.

potential $\tilde{A}_a = A_a - \nabla_a \Lambda = 0$. In general however, $\nabla_a \Lambda$ need not exist. Thus $\tilde{\psi}(x)$ though continuous, need not be differentiable and it may therefore not satisfy a differential equation. Nevertheless it is clear from (3.10) that there exists a (1−1) correspondence between the solutions of the usual wave equation in any given gauge and the gauge-independent $\tilde{\psi}(x)$ defined here, at least in the absence of magnetic monopoles.

The gauge transformation $A_a \to A_a - \nabla_a \alpha(x)$ results in the local gauge transformation $\psi(x) \to \psi(x) e^{ie\alpha/\hbar(x)}$. But $\tilde{\psi}(x)$ is unchanged under such a transformation and in this sense $\tilde{\psi}$ is gauge independent. However if the field of reference paths $\{\gamma(x)\}$ of the reference point x_0 is changed, the new wave function $\tilde{\psi}'(x)$ is also related to $\psi(x)$ by an equation of the form (3.10). Hence $\tilde{\psi}'(x) = e^{i\beta(x)}\tilde{\psi}(x)$ for some continuous function $\beta(x)$, which need not be differentiable.

If F^{ab} is due to magnetic charges in addition to possible electric charges, then the present formalism is inequivalent to Feynman's formalism. In this case A_a must have a line of singularity through each magnetic charge. So Feynman's prescription for the propagator (3.7) requires at least that none of the paths γ_r contain a connected portion of a line of singularity. This requirement, for which there is no physical justification, is not needed in the present formalism. But on the other hand, the present formalism raises the difficult problem of showing that the path integral in (3.5) can be performed when magnetic monopoles are present, which is necessary in order that it be useful in this case. This problem will not be attempted in this paper.

As already mentioned, to write (3.5) we needed the correspondence principle (2.2) and the Lorentz force law (2.14), but did not require a knowledge of the classical Lagrangian for the interaction between the particle and the electromagnetic field. This suggests that (2.2) or a suitable generalization of it may be useful in quantizing theories that have a classical equation of motion, but no Lagrangian or Hamiltonian.

Finally, consider the case of a particle moving in the electromagnetic field of another particle, with both particles having electric and magnetic charges e_1, e_2, g_1, and g_2, respectively. If the classical equation of motion for the first particle is taken to be the following generalization of (2.14),

$$\frac{Dp^a}{Ds} = \frac{e_1}{c} F^{ab} v_b + \frac{g_1}{c} {}^*F^{ab} v_b \tag{3.11}$$

where ${}^*F^{ab} = \frac{1}{2}\varepsilon^{abcd}F_{cd}$, then by (2.2), the phase shift for an infinitesimal interference loop is

$$\Delta\phi = \frac{e_1}{2\hbar} F_{ab}\sigma^{ab} + \frac{g_1}{2\hbar} {}^*F_{ab}\sigma^{ab} \tag{3.12}$$

Finally, assumption A, on using an argument similar to that which gave (3.4), yields

$$e_1 g_2 - e_2 g_1 = \frac{n\hbar}{2}, \qquad n = 0, \pm 1, \ldots \tag{3.13}$$

which is the generalization of (3.4) for the more general case being considered here. (3.13) can also be regarded as the duality-invariant way of expressing (3.4), the absence of magnetic monopoles corresponding to the ratio e/g being the same for all particles. (See, for instance, Jackson, 1975). Also (3.5), for a particle with electric charge e and magnetic charge g, is given by

$$\tilde{K}(x,x') = \sum_r \exp\left(\frac{i}{2\hbar} \int_{S_r} (eF_{ab} + g^* F_{ab})\, d\sigma^{ab} \right) \kappa_r \tag{3.14}$$

It is clear from the above discussion that (3.5) can be generalized to any gauge field corresponding to an *Abelian* gauge group, since assumption A, formulated at the beginning of this section, is valid in this case. We now consider cases for which assumption A is not applicable or not valid: First of all, in a multiply connected region R, it is not possible to find a surface spanned by every closed curve γ, which lies entirely within R. For instance, in the original experiment discussed by Aharonov and Bohm (1959), we may choose R to be the region outside the cylinder containing the magnetic field. If one is not permitted to look inside the cylinder than one cannot obtain the phase shift for this experiment by the use of assumption A. Hence it is clear that in a multiply connected region we must replace $\exp(ie/2\hbar \int_{S_r} F_{ab}\, d\sigma^{ab})$ in (3.5) by $\exp(ie/2\hbar \oint_{\Gamma_r} A_a\, dx^a)$, where Γ_r is the closed curve formed from γ_1 and γ_r.

Consider now, as the external field, a gauge field corresponding to an arbitrary gauge group G having generators T_i. The classical equation for a particle in this field is

$$\frac{DP^a}{Ds} = \tau_i F^{ia}{}_b v^b \tag{3.15}$$

where F^i_{ab} is the gauge field strength and τ_i are the gauge parameters of the particle (e.g., isospin). This is the generalization to curved space-time and arbitrary gauge field of the equation, considered by Wong (1970), for the $SU(2)$ gauge field. It follows from (3.15) and the correspondence principle (2.2), (2.3) that the phase shift due to the field infinitesimally is

$$\Delta\phi = \frac{c}{2\hbar} \tau_i F^i{}_{ab}\, d\sigma^{ab} \tag{3.16}$$

Now we cannot use the analog of assumption A to obtain the phase shift for a finite interference loop, when G is non-Abelian, because such an assumption is not valid for this case. This is because, in this case, $e^{i\Delta\phi}$ for a given finite interference loop *cannot* be written as $\exp(i\Sigma_\alpha \Delta\phi_\alpha)$, where $\Delta\phi_\alpha$ is defined above (3.1). But it is shown elsewhere (Anandan, 1979) that the phase shift in interference due to an arbitrary gauge field, under certain conditions, is determined by an operator of the form

$$O\left(\exp\left(i\oint_\gamma A_a^i T_i dx^a\right)\right) \tag{3.17}$$

where A_a^i is the gauge potential, O denotes path ordering, and the integral is along a closed curve γ beginning and ending at a point in the interference region. It is therefore reasonable to suppose that the generalization of (3.5) to an arbitrary gauge field is

$$K(x,x')\sim\sum_r O\exp\left(i\oint_{\Gamma_r} A_a^i \mathbf{T}_i dx^a\right)\kappa_r \tag{3.18}$$

where Γ_r is the closed curve formed from γ_1 and γ_r, κ_r (an operator) is associated with the curve γ_r (joining x' to x) in the absence of the gauge field, and $\mathbf{T}_i$ provides the representation of interest of T_i. The evolution of the wave function is determined as before by (3.6).

It was pointed out by Wu and Yang (1975) that the field strength of an $SU(2)$ gauge field does not have all the gauge-invariant information of the field, in general, because it is possible to have two gauge-inequivalent $SU(2)$ gauge potentials which have the same field strength. This raises the question of whether the operators (3.17) associated with closed loops γ, which determine the phase shift, contain all the gauge-invariant information about the gauge field. We answer this question affirmatively in the Appendix by showing that *two gauge potentials corresponding to an arbitrary gauge group are related by an active gauge transformation if and only if the path-ordered operators (3.17) for the two gauge potentials, associated with each curve γ beginning and ending at an arbitrary (but fixed) point, are related by a constant similarity gauge transformation*. It follows that, when G is non-Abelian, in general it is not possible to go from the classical equation of motion (3.15) to the quantum theory as we did in the Abelian case, even in a simply connected region, because (3.15) depends on the field strength, whereas (3.17), which is needed for the quantum theory, has more information, in general. These remarks also show the importance of interference in operationally defining a gauge field: From the response of the energy-momentum of a classical test particle to a gauge field we can determine, via (3.15), only the field strength, which, however, does not contain all the gauge-invariant information about the field, in general. The "isospin" τ_i is covariantly constant along the classical trajectory. So from its motion, the adjoint representation of the operators (3.17) can be determined. But in general, this also does not have all the gauge invariant information of the gauge field. It is interference that enables us to determine the holonomy group consisting of elements of the form (3.17), which contains all the gauge-invariant information about the field, according to the above theorem.

ACKNOWLEDGMENTS

It is a pleasure to thank Paul Green for a useful discussion concerning the theorem in the Appendix.

APPENDIX: GAUGE FIELDS, GRAVITY, AND HOLONOMY

In this section we shall state and prove a theorem which contains, as a special case, the theorem stated in the last paragraph of Section 3. It is convenient, for this purpose, to regard a gauge field as a connection in a principal fiber bundle over space-time. It will be assumed that the reader is familiar with this description,[8] which also has the advantage that the theorem, when stated in this language, is readily seen to apply to the gravitational field connection as well.

Let $P(M,G)$ and $\tilde{P}(\tilde{M},\tilde{G})$ be two principal fiber bundles over base manifolds M, $\tilde{M}$ with structure groups G, $\tilde{G}$ whose right actions are denoted by $R_a(a \in G)$ and $\tilde{R}_b(b \in \tilde{G})$. A homomorphism f: $P(M,G) \rightarrow \tilde{P}(\tilde{M},\tilde{G})$ consists of a differentiable map f': $P \rightarrow \tilde{P}$ and a homomorphism f'': $G \rightarrow \tilde{G}$ such that $f'R_a = R_{f''(a)}f'$ for every $a \in G$. A homomorphism f is an imbedding if f' is an imbedding and f'' is (1–1). For simplicity, f' and f'' are sometimes denoted by the same letter f.

> *Theorem.* Let $P(M,G)$ and $\tilde{P}(\tilde{M},\tilde{G})$ with projection maps π and $\tilde{\pi}$ have connections Γ and $\tilde{\Gamma}$, respectively. Then there exists an imbedding f: $P(M,G) \rightarrow \tilde{P}(\tilde{M},\tilde{G})$, which maps Γ into $\tilde{\Gamma}$ if and only if there exists a differentiable map h: $M \rightarrow \tilde{M}$ and for any $x \in M$ there exists a (1–1) differentiable map g: $\pi^{-1}(x) \rightarrow \tilde{\pi}^{-1}(h(x))$ and a (1–1) homomorphism g': $G \rightarrow \tilde{G}$ such that (i) $gR_a = \tilde{R}_{g'(a)}g$ for every $a \in G$ and (ii) for every curve γ_x, beginning and ending at x, $\gamma_x = g^{-1}\tilde{\gamma}_{h(x)}g$, where γ_x and $\tilde{\gamma}_{h(x)}$ are, respectively, the holononomy transformations in $P(M,G)$ and $\tilde{P}(\tilde{M},\tilde{G})$ associated with the connections Γ and $\tilde{\Gamma}$, corresponding to the closed curves γ_x and $\tilde{\gamma}_{h(x)} \equiv h(\gamma_x)$[$\tilde{\gamma}_{h(x)}$ is the image of γ_x under h and it is understood that the domain of g^{-1} is $g(\pi^{-1}(x))$].

[8]See, for instance, Kobayashi and Nomizu (1969).

Proof. Suppose there exists an imbedding f: $P(M,G)\to\tilde{P}(\tilde{M},\tilde{G})$ which maps Γ into $\tilde{\Gamma}$. Define h: $M\to\tilde{M}$ to be the map induced on M by f. Consider any $x\in M$ and a curve γ beginning and ending at x. Define $g=f|_{\pi^{-1}(x)}$. Now f maps horizontal curves in $P(M,G)$ into horizontal curves in $\tilde{P}(\tilde{M},\tilde{G})$. Since the holonomy transformations γ_x or $\tilde{\gamma}_{h(x)}$ are defined by horizontal lifts of γ_x or $\tilde{\gamma}_{h(x)}$ with respect to Γ or $\tilde{\Gamma}$, respectively, it follows immediately that $\gamma_x=g^{-1}\tilde{\gamma}_{h(x)}g$.

Conversely, suppose that there exist differentiable maps h: $M\to\tilde{M}$, g: $\pi^{-1}(x)\to\tilde{\pi}^{-1}(h(x))$, for some $x\in M$, and a (1–1) homomorphism g': $G\to\tilde{G}$ such that $gR_a=\tilde{R}_{g'(a)}g$ for every $a\in G$ and $\gamma_x=g^{-1}\tilde{\gamma}_{h(x)}g$ for every curve γ beginning and ending at x, where $\gamma_{h(x)}\equiv h(\gamma_x)$. Denote $h(x)$ by $\tilde{x}(x\in M)$. Now define f': $P\to\tilde{P}$ in the following way: For every $y\in M$, let γ_{xy} be a curve joining y to x and $\tilde{\gamma}_{\tilde{x}\tilde{y}}$ be the image of this curve under the map h. Let γ_{xy} and $\tilde{\gamma}_{\tilde{x}\tilde{y}}$ denote the parallel displacement along γ_{xy} and $\tilde{\gamma}_{\tilde{x}\tilde{y}}$ with respect to the connections Γ and $\tilde{\Gamma}$, respectively. Define f' by $f'|_{\pi^{-1}(y)}=\tilde{\gamma}_{\tilde{x}\tilde{y}}^{-1}g\gamma_{xy}$ $(y\in M)$. It is easily verified that $f'|_{\pi^{-1}(y)}$ is independent of the chosen curve γ_{xy} joining y to x so that f': $P\to\tilde{P}$ is well defined. Also

$$f'R_a|_{\pi^{-1}(y)}=\tilde{\gamma}_{\tilde{x}\tilde{y}}^{-1}g\gamma_{xy}R_a|_{\pi^{-1}(y)}=\tilde{\gamma}_{\tilde{x}\tilde{y}}^{-1}gR_a\gamma_{xy}|_{\pi^{-1}(y)}$$
$$=\gamma_{\tilde{x}\tilde{y}}^{-1}\tilde{R}_{g'(a)}g\gamma_{xy}|_{\pi^{-1}(y)}=\tilde{R}_{g'(a)}f'|_{\pi^{-1}(y)}$$

for every $y\in M$ and $a\in G$. Therefore, on defining $f''=g'$, $f'R_a=\tilde{R}_{f''(a)}f'$. The maps f' and f'' therefore constitute an imbedding f: $P(M,G)\to\tilde{P}(\tilde{M},\tilde{G})$.

Finally, let γ_{zy} be an arbitrary curve in M joining $y\in M$ to $z\in M$. Then by supposition, $g^{-1}\tilde{\gamma}_{\tilde{x}\tilde{z}}\tilde{\gamma}_{\tilde{z}\tilde{y}}\tilde{\gamma}_{\tilde{x}\tilde{y}}^{-1}g=\gamma_{xz}\gamma_{zy}\gamma_{xy}^{-1}$, i.e., $\tilde{\gamma}_{\tilde{z}\tilde{y}}\tilde{\gamma}_{\tilde{x}\tilde{y}}^{-1}g\gamma_{xy}=\tilde{\gamma}_{\tilde{x}\tilde{z}}^{-1}g\gamma_{xz}\gamma_{zy}$, which is the same as $\tilde{\gamma}_{\tilde{z}\tilde{y}}f'=f'\gamma_{zy}$. Therefore f maps horizontal curves of P into horizontal curves of $\tilde{P}$. Hence f maps Γ into $\tilde{\Gamma}$. This completes the proof. ■

Consider now the special case when $P=\tilde{P}$, $M=\tilde{M}$, $G=\tilde{G}$, and $\pi=\tilde{\pi}$. Then in the above theorem f: $P(M,G)\to P(M,G)$ is an isomorphism. If in addition f preserves each fiber then h: $M\to M$ is the identity map, g and g' are automorphisms, and the curves γ_x and $\tilde{\gamma}_{h(x)}$ are the same. Also if M is space-time then Γ and $\tilde{\Gamma}$ may be regarded as gauge field connections. Then f represents an active gauge transformation while g is an active gauge transformation at a point. Under these conditions, the above theorem is then the same as the theorem stated at the end of Section 3.[9] Also if

[9]As a simple illustration of this theorem consider the Wu–Yang copies referred to in the last paragraph of Section 3. It is easy to verify that these two potentials have as their holonomy groups $SU(2)$ and $U(1)$. It then follows trivially from the above theorem that they must therefore be gauge inequivalent.

$P(M, G)$ is the bundle of frames on M then Γ has the interpretation of the gravitational field connection. A necessary and sufficient condition that there exists a Lorentzian metric compatible with this connection (i.e., the covariant derivative of the metric with respect to this connection is zero) is that the holonomy group is a subgroup of the Lorentz group.

REFERENCES

Aharonov, Y., and Bohm, D. (1959). *Physical Review*, **115**, 484.

Anandan, J. (1977). *Physical Review* D, **15**, 1448.

Anandan, J. (1979). *Il Nuovo Cimento*, **53A**, 221.

Anandan, J. (1980). *Foundations of Physics* (to be published).

Cabibbo, N., and Ferrari, E. (1962). *Il Nuovo Cimento*, **23**, 1147.

Chambers, R. G. (1960). *Physical Review Letters*, **5**, 3.

Dirac, P. A. M. (1931). *Proceedings of Royal Society, London Series A*, **133**, 60.

Dirac, P. A. M. (1948). *Physical Review*, **74**, 817.

Ehlers, J., Pirani, F. A. E., and Schild, A. (1972). *General Relativity* L. O'Raifeartaigh, ed. Clarendon Press, Oxford.

Feynman, R. P. (1948). *Reviews of Modern Physics*, **20**, 367.

Feynman, R. P. (1964). *The Feynman Lectures on Physics*, Vol. II, Sec. 15-5, Vol. III, Sec. 7-4. Addison-Wesley, Reading, Massachusetts.

Jackson, J. D. (1975). *Classical Electrodynamics*, 2nd ed., 6-12, p. 252. John Wiley, New York.

Kobayashi, S., and Nomizu, K. (1969). *Foundations of Differential Geometry*, Interscience, New York.

Marzke, R. F., and Wheeler, J. A. (1964). *Gravitation and Relativity*, edited by H. Y. Chiu and W. F. Hoffmann, eds., Benjamin, New York.

Weyl, H. (1921). *Nachr Gottingen*, **99**.

Wong, S. K. (1970). *Il Nuovo Cimento*, **LXV A**, 689.

Wu, T. T., and Yang, C. N. (1975). *Physical Review D* **12**, 3843, 3845.