International Journal of Theoretical Physics, Vol. 41, No. 2, February 2002 (© 2002)

Symmetries, Quantum Geometry, and the Fundamental Interactions

Jeeva Anandan[1]

Received April 25, 2001

A generalized Noether's theorem and the operational determination of a physical geometry in quantum physics are used to motivate a quantum geometry consisting of relations between quantum states that are defined by a universal group. Making these relations dynamical implies the nonlocal effect of the fundamental interactions on the wave function, as in the Aharonov–Bohm effect and its generalizations to non-Abelian gauge fields and gravity. The usual space–time geometry is obtained as the classical limit of this quantum geometry using the quantum-state space metric.

1. INTRODUCTION

The space–time geometry that is commonly used today arose from classical physics. An interesting question is what geometry is appropriate for quantum physics. It was suggested that the universal symmetry group elements that act on all Hilbert spaces may be appropriate for constructing a physical geometry for quantum theory (Anandan, 1980). I also proposed the systematic study of all the fundamental interactions *operationally* from their effects on quantum interference (Anandan, 1979). The purpose of this paper is to attempt to bring together these two approaches. The modular variables introduced by Aharonov *et al.* (1969) will play a useful role in this.

In Section 2, I shall review Noether's theorem and its converse in a generalized form in which the conserved quantities are elements of a group and not the generators of this group as usually stated. This will suggest a quantum geometry by relations defined by the universal group elements, which constitute the symmetry group of physics, that act on all Hilbert spaces, as discussed in Section 3. The classical limit of this geometry will be obtained in Section 4 as the usual space–time geometry from the quantum-state metric in Hilbert spaces and the universality of the action of the translational group in every Hilbert space. In Section 5, the

[1] Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208; e-mail: jeeva@sc.edu.

0020-7748/02/0200-0199/0

nonlocality of fundamental interactions in quantum physics implied by this approach, as shown physically by the Aharonov–Bohm effect (Aharonov and Bohm, 1959) and its generalizations (Anandan, 1979; Wisnievesky and Aharonov, 1967), will be studied. The study of the gravitational Aharonov–Bohm effect around a cosmic string in particular suggests that the use of universal group elements as quantum distances may be appropriate.

2. SOME REFLECTIONS ON NOETHER'S THEOREM

The usual statement of Noether's theorem is that for every continuous symmetry of the equations of motion (determined by the Lagrangian or Hamiltonian) there exists a conserved quantity. Although this is easy to prove, the meaning of this theorem is more readily evident in the Hamiltonian than in the Lagrangian formulation. A symmetry of the equations of motion or time evolution is a transformation s such that in any experiment if s is applied to the initial state of the physical objects (fields, particles, etc.) participating in the experiment then the final state of the transformed experiment must be the same as s applied to the final state of the original experiment. For example, spatial translational symmetry implies that if the apparatus is translated to a new spatial location then the same experiment should give the same result.

Suppose U is the time-evolution operator, and ψ_i and ψ_f are the initial and final states, i.e. $\psi_f = U\psi_i$. Then the above definition of s being a symmetry of the time evolution is

$$s\psi_f = Us\psi_i$$

for every initial state ψ_i. This is equivalent to the commutator

$$[U, s] = 0. \tag{1}$$

But (1) states also that s is conserved during the time evolution. Therefore, the statements that s is a symmetry and that s is conserved are the *same* statement (1), and there is nothing to prove!

Now suppose that there is a continuous symmetry generated by Q. Then (1) is satisfied with $s = \exp(iQq)$ for all q, and therefore

$$[U, Q] = 0. \tag{2}$$

Hence, Q is conserved, which is Noether's theorem. Furthermore, if the Hamiltonian H is independent of time t, as it is for an isolated system, then $U = \exp(-\frac{i}{\hbar}Ht)$. If (2) is valid for all t, then

$$[H, Q] = 0. \tag{3}$$

The above results may be extended to classical physics by turning the above commutators into Poisson brackets in classical phase space. These classical results may be regarded as the classical limit of the quantum results by recognizing that

the symplectic structure that gives the Poisson brackets are the classical limit (Anandan, 1990, 1991) of a symplectic structure in quantum theory (Kibble, 1979) that gives the commutators.

But the conservation of s that follows from (1) is more general than the usual form of Noether's theorem. There are at least two situations in which (1) is valid but there are no corresponding (2) or (3). First, as is well known, in both classical and quantum physics, s may be a discrete symmetry instead of a continuous symmetry. For example, s may be parity, which is a symmetry and therefore conserved for all interactions except the weak interaction, as far as we know. Another example is that (1) is satisfied for $s = \exp(i\,Qq_k)$ for a discrete set of values q_k only. Second, in quantum physics the mean value of $s = \exp(i\,Qq)$ has more information than the mean value of all the moments of Q, namely Q^n where n is any positive integer (Aharonov *et al.*, 1969). This is unlike in classical physics where the mean value of a transformation generated by Q may be obtained using the mean values of all the Q^n. Both these situations will be considered in Section 5.

Since in (1) U and s occur symmetrically, it follows that the converse of the generalized Noether's theorem is also true: *A transformation s that is conserved must be a symmetry of the equations of motion*. The usual view is that U is more fundamental than s because U is determined by the dynamical laws, which are regarded as primary, whereas the symmetries such as s are obtained secondarily as the symmetries of these laws. But the concise form (1) of the connection between the dynamical laws and symmetries, in which U and s are on an equivalent footing, suggest that we may equally turn the usual view around and regard the transformations $\{s\}$ as primary and U as derived from them to satisfy (1) so that $\{s\}$ are the symmetries (Anandan, 1999). The possibility of regarding symmetries as fundamental relations between quantum states by associating them with a quantum geometry will be explored in the next section.

3. QUANTUM GEOMETRY

The concept of space originates from our common experience of translating objects and from the possible states they can occupy. If we translate a cup, for example, in various possible ways, classically we may say that the different configurations or states of the cup are "immersed" in "space." This space is *universal* in the sense that it is regarded as independent of the objects "contained" in it.

But quantum mechanically it is not clear what is meant by the cup being "immersed" in space. The cup consists of electrons, protons, and neutrons (or the quarks and gluons that make up the protons and neutrons), and the states of these particles belong to the corresponding Hilbert spaces different from the physical space or the phase space of classical physics. The translation of a cup therefore needs to be represented by the corresponding translation operators that act on these Hilbert spaces. The fact that all the particles constituting the cup move together in

some approximate sense suggests the introduction of universal translation group elements that are represented by operators that act on each Hilbert space. It is this *universality* of the translation group that gives us the concept of "space" that is independent of the particular system that partakes in it.

Also, it is well known that we cannot operationally determine the metric in space–time, or even the points of space–time, using quantum probes (Wigner, 1967) because of the uncertainty principle. If one tries to obtain the space–time geometry using a clock and radar light signals, which is possible in classical physics (Synge, 1960), the uncertainty in the measurement of time in quantum physics is approximately equal to $\hbar/\Delta E$, where ΔE is the uncertainty in the energy of the clock. If we try to decrease this uncertainty by increasing ΔE, then this increases the uncertainty in the geometry of space–time due to the uncertainty in the gravitational field of the clock. The total uncertainty in the measurement of space–time distances is then $\hbar c/\Delta E + 2G\Delta E/c^4$. The minimum value of this uncertainty as ΔE is varied is approximately the Planck length $= \sqrt{G\hbar/c^3}$. Hence, space–time geometry is only approximately valid in quantum theory, with an uncertainty of the order of Planck length.

However, a geometry for quantum theory may be defined by relations determined by a universal group S, which generalizes the above translation group. This is universal in the sense that the same S has a representation in each Hilbert space. But S may have subgroups that may have trivial representations in some Hilbert spaces but not in others. An object may be displaced by any $s \in S$, which means that s acts on each of the Hilbert spaces of the particles or fields constituting that object through the corresponding representation of S. Each ψ in each of these Hilbert spaces is mapped to a corresponding ψ_s by this action of s, and the group element s that determines the *relation* between ψ and ψ_s is independent of the Hilbert space and is therefore universal. In the example of a cup considered above, s is an element of the translation group, ψ and ψ_s are the states of each particle constituting the cup before and after the translation, and the relation between each such pair is universal in the sense that the entire cup has undergone this translation, or any other object that could take the place of the cup. This quantum geometry cannot be subject to the above criticisms of the space–time geometry because the action of S on each Hilbert space is not subject to any uncertainty.

It is reasonable to require that this geometry is preserved in time in the absence of interactions. Then each $s \in S$ is conserved and the converse of Noether's theorem stated in the last paragraph of the previous section implies that s is also a symmetry of the time evolution. Since the evolution equations are now determined by the standard model, S may be the symmetry group of the present-day standard model, namely $P \times U(1) \times SU(2) \times SU(3)$, where P is the Poincare group. But if the standard model is superseded by new physics that has a different symmetry group then S should be this new group, and the above statements would all be unaffected.

As an illustration of the geometrical relations proposed here, consider the experimentally known quantization of electric charge: all known charges are integral multiples of the fundamental charge e_0. An aspect of this is that the magnitudes of the charges of the electron and the proton are experimentally known to be equal to an amazing precision. To obtain charge quantization, take s above to be an arbitrary element of the electromagnetic $U(1)$ group, which is a subgroup of S. This universal $U(1)$ group is a circle parametrized by Λ, say, that varies from 0 to Λ_0 so that 0 and Λ_0 represent the same point on this group, chosen to be the identity. Since $U(1)$ is Abelian, it has only one-dimensional representations. Hence the action of $s(\Lambda)$ on an arbitrary state gives

$$\psi_s = \exp(i\,Q\Lambda)\psi, \tag{4}$$

where Q corresponds to a particular representation of $U(1)$ in the Hilbert space in which ψ belongs to. But since $s(\Lambda_0) = s(0)$, which is because of the compactness of the $U(1)$ group, $\exp(i\,Q\Lambda_0) = 1$ for all representations (Yang, 1970). Hence, $Q\Lambda_0 = 2\pi n$ or

$$Q = n\frac{2\pi}{\Lambda_0}, \tag{5}$$

where n is an integer.

To interpret Q, consider the physical implementation of the transformation s. This may be done by sending each of the particles through the same electromagnetic field with four-vector potential A_μ in a particular gauge so that the effect of the electromagnetic field alone on the particle is given by

$$\psi_s = \exp\left(-i\frac{q}{\hbar c}\int A_\mu(x)\,dx^\mu\right)\psi, \tag{6}$$

which is a $U(1)$ transformation. Indeed, the statement that the electromagnetic field is a $U(1)$ gauge field may be taken to mean that it is physically possible to implement a $U(1)$ gauge transformation using the electromagnetic field in this way. Then q has the interpretation of the electric charge. Comparing (6) with (4), we may take Λ to be $\int A_\mu\,dx^\mu$, in which case $Q = q/\hbar c$. Hence, from (5),

$$q = n\,e_0, \tag{7}$$

where $e_0 = 2\pi\hbar c/\Lambda_0$ is a universal constant that is determined experimentally to be $\frac{1}{3}e$, where e is the charge of the electron. The exact equality of the magnitudes of the charges of the electron and the proton may now be understood as due to them belonging to representations corresponding to $n = 3$ and $n = -3$, respectively.

The above argument also provides a reason for the introduction of Planck's constant, which is purely geometrical. The exponent in (6) must be dimensionless because the expansion of the exponential has all powers of the exponent. Now,

$\frac{q}{c}\int A_\mu\, dx^\mu$ is meaningful in classical physics. But to turn it into a dimensionless quantity, it is necessary to introduce a new scale, which is provided by $\hbar$. From the present point of view, this is needed in order to physically implement the $U(1)$ group elements that define relations between states that are part of the quantum geometry. Also, from (7), q is proportional to e, and A_μ is also proportional to e because the charges that generate A_μ via Maxwell's equations are proportional to e. Hence, the exponent in (6) is proportional to the fine-structure constant $e^2/\hbar c$. This argument may be extended to gauge fields in general, and dimensionless coupling constants are obtained for all of them. From now on, units in which $c = 1$ will be used, and the metric convention is $(+, -, -, -)$.

The relation defined by (6) is not gauge-invariant. Hence, it cannot be used to define an invariant geometrical "distance." Consider again the translation of a cup that may be performed by acting on all the quantum states of the particles constituting the cup by a universal group element $\exp(-\frac{i}{\hbar}\hat{p}\ell)$, where $\hat{p}$ is a generator of translation. The action of this group element on a wave function is also not gauge-invariant. But we may combine the two transformations to define the gauge-covariant transformation $\psi_\ell(x) = f_\ell(x)\psi(x)$, where x stands for x^μ or equivalently $(\mathbf{x}, t)$, and[2]

$$f_\ell(x) = \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}\right)\exp\left\{i\frac{q}{\hbar}\int_{\mathbf{x}}^{\mathbf{x}+\boldsymbol{\ell}}\mathbf{A}(\mathbf{y}, t)\cdot d\mathbf{y}\right\} \tag{8}$$

where at present the integral is taken along the straight line joining $x = (\mathbf{x}, t)$ and $(\mathbf{x} + \boldsymbol{\ell}, t)$ for simplicity. Then, clearly,

$$f_\ell(x) = \exp\left\{i\frac{q}{\hbar}\int_{\mathbf{x}-\boldsymbol{\ell}}^{\mathbf{x}}\mathbf{A}(\mathbf{y}, t)\cdot d\mathbf{y}\right\}\exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}\right). \tag{9}$$

It is easy to show, using (8) and (9), that the set of operators $\{f_\ell \mid \ell \in \mathcal{R}^3\}$ is a group under multiplication.

From (8), under a gauge transformation, $\psi'(x) = u(x)\psi(x)$, where $u(x) = \exp\{i\frac{q}{\hbar}\Lambda(x)\}$, and $A'_\mu(x) = A_\mu(x) - \partial_\mu\Lambda(x)$, f_ℓ transforms to

$$f'_\ell(x) = \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}\right)u(\mathbf{x}+\boldsymbol{\ell}, t)\exp\left\{i\frac{q}{\hbar}\int_{\mathbf{x}}^{\mathbf{x}+\boldsymbol{\ell}}\mathbf{A}(\mathbf{y}, t)\cdot d\mathbf{y}\right\}u(\mathbf{x}, t)$$

On using

$$\exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}\right)u(\mathbf{x}+\boldsymbol{\ell}, t) = u(\mathbf{x}, t)\exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}\right) \tag{10}$$

[2] The action of $\exp(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}) = \exp(-\boldsymbol{\ell}\cdot\nabla)$ on an analytic wave function ψ may be understood as the usual Taylor expansion of ψ. If ψ is not analytic, then $\exp(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell})$ acts on the momentum space wave function $\tilde{\psi}$ according to $\tilde{\psi}(\mathbf{p}, t) \to \exp(-\frac{i}{\hbar}\mathbf{p}\cdot\boldsymbol{\ell})\tilde{\psi}(\mathbf{p}, t)$. In either case, we obtain $\exp(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell})\psi(\mathbf{x}, t) = \psi(\mathbf{x} - \boldsymbol{\ell}, t)$.

it follows that

$$f'_\ell(x) = u(\mathbf{x}, t)\, \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\ell\right)\, \exp\left\{i\frac{q}{\hbar}\int_{\mathbf{x}}^{\mathbf{x}+\ell}\mathbf{A}(\mathbf{y}, t)\cdot d\mathbf{y}\right\} u^\dagger(\mathbf{x}, t)$$
$$= u(x)\, f_\ell(x) u^\dagger(x). \tag{11}$$

Hence, f_ℓ acts gauge-covariantly on the Hilbert space.

f_ℓ may also shown to be gauge-covariant from the fact that

$$\langle\psi|\, f_\ell|\psi\rangle = \left\langle \exp\left(\frac{i}{\hbar}\mathbf{p}\cdot\ell\right)\psi \,\middle|\, \exp\left\{i\frac{q}{\hbar}\int_{\mathbf{x}}^{\mathbf{x}+\ell}\mathbf{A}(\mathbf{y}, t)\cdot d\mathbf{y}\right\}|\psi\rangle\right.$$
$$= \int d^3x\, \psi^\dagger(\mathbf{x}+\ell, t)\, \exp\left\{i\frac{q}{\hbar}\int_{\mathbf{x}}^{\mathbf{x}+\ell}\mathbf{A}(\mathbf{y}, t)\cdot d\mathbf{y}\right\}\psi(\mathbf{x}, t) \tag{12}$$

is gauge-invariant because the integrand is gauge-invariant (Anandan, 1986). It also follows from (12) that the operator (8) is observable. For example, in the Josephson effect, where the current depends on the gauge-invariant phase difference across the junction, if ℓ is chosen to be the vector across the junction then f_ℓ is observable from the current (Anandan, 1986).

A gauge may be chosen, even if the field strength is nonvanishing—so that the component of $\mathbf{A}$ in the direction of ℓ is zero. Then in this gauge $f_\ell = \exp(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\ell)$, which is conserved for an isolated system. Since f_ℓ is gauge-covariant, it follows that (8) is conserved for an isolated system in every gauge. f_ℓ is a gauge-covariant generalization of the modular momentum (Aharonov *et al.*, 1969), and may be called the modular kinetic momentum. When two systems interact, there would be an exchange of modular kinetic momentum. Indeed, this may be regarded as the definition of two systems interacting (Anandan, 1999). As will be seen in Section 5, this exchange may happen even when there are no forces between the two systems, which makes the latter interaction more general than the usual interaction via forces. If $\tilde{\psi}$ is the momentum space wave function of ψ, then clearly

$$\langle\psi|\exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\ell\right)|\psi\rangle = \int d^3p|\tilde{\psi}(\mathbf{p}, t)|^2\, \exp\left(-\frac{i}{\hbar}\mathbf{p}\cdot\ell\right). \tag{13}$$

Hence, the change in $\langle\psi|\exp(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\ell)|\psi\rangle$ results in a change in the distribution of the probability density $|\tilde{\psi}(\mathbf{p}, t)|^2$ of momentum $\mathbf{p}$ (Aharonov, private communication). Since in the above gauge the gauge-invariant $\langle\psi|\, f_\ell|\psi\rangle$ is the same as (13), it follows that, in an arbitrary gauge, an interaction results in a change in the distribution of the gauge-invariant kinetic momentum $\mathbf{p} - q\mathbf{A}$.

Similarly, the modular energy (Aharonov *et al.*, 1969) may be generalized gauge-covariantly as follows. Define the transformation on the Hilbert space f_τ

by $\psi_\tau(x) = f_\tau(x)\psi(x)$, where

$$f_\tau(x) = \left\{T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H\,dt\right)\right\}^\dagger\ \exp\left\{-i\frac{q}{\hbar}\int_t^{t+\tau} A_0(x)\,dt\right\} \tag{14}$$

and T denotes time ordering. To prove that f_τ is gauge-covariant under the above gauge transformation, note first that for a solution $\psi(\mathbf{x}, t)$ of Schrödinger's equation,

$$\psi(\mathbf{x}, t+\tau) = T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H\,dt\right)\psi(\mathbf{x}, t). \tag{15}$$

Therefore, under the gauge transformation $\psi'(\mathbf{x}, t) = u(\mathbf{x}, t)\psi(\mathbf{x}, t)$,

$$\psi'(\mathbf{x}, t+\tau) \equiv u(\mathbf{x}, t+\tau)\psi(\mathbf{x}, t+\tau) = u(\mathbf{x}, t+\tau) \times T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H\,dt\right)u^\dagger(\mathbf{x}, t)\psi'(\mathbf{x}, t)$$

using (15). Hence,

$$u(\mathbf{x}, t+\tau)T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H\,dt\right)u^\dagger(\mathbf{x}, t) = T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H'\,dt\right),$$

where H' is the gauge-transformed Hamiltonian. The last equation is equivalent to

$$\left\{T\exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H'\,dt\right)\right\}^\dagger u(\mathbf{x}, t+\tau) = u(\mathbf{x}, t) \times \left\{T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H\,dt\right)\right\}^\dagger. \tag{16}$$

Under this gauge transformation, $f_\tau(x)$ transforms to

$$\begin{aligned} f'_\tau(x) &\equiv \left\{T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H'\,dt\right)\right\}^\dagger \exp\left\{-i\frac{q}{\hbar}\int_t^{t+\tau} A'_0(x)\,dt\right\} \\ &= \left\{T\ \exp\left(-\frac{i}{\hbar}\int_t^{t+\tau} H'\,dt\right)\right\}^\dagger u(\mathbf{x}, t+\tau) \\ &\quad \times \exp\left\{-i\frac{q}{\hbar}\int_t^{t+\tau} A_0(x)\,dt\right\}u^\dagger(\mathbf{x}, t) \\ &= u(\mathbf{x}, t)\,f_\tau u^\dagger(\mathbf{x}, t) \end{aligned} \tag{17}$$

on using (16). Hence, f_τ acts gauge-covariantly on the Hilbert space.

More generally, for an arbitrary gauge field the above arguments hold with eA_μ replaced by $g_0 A_\mu^k T_k$, where T_k generate the gauge group, and $u(x)$ is the corresponding local gauge transformation. The gauge field exponential (parallel transport) operator will be generalized to be along an arbitrary piecewise-differentiable curve $\hat{\gamma}$ in $\mathcal{R}^4$. Path ordering is necessary because $\{T_k\}$ do not commute in general. The energy–momentum operator $\hat{p}_\mu$ is defined by $\hat{p}_0 = H$, $\hat{p}_i = i\hbar \frac{\partial}{\partial x^i}$. Now define the transformation g_γ on the Hilbert space by $\psi_\gamma(x) = g_\gamma(x)\psi(x)$, where

$$g_\gamma(x) = P\ \exp\left(-\frac{i}{\hbar}\int_{\bar{\gamma}} \hat{p}_\mu\, dy^\mu\right) P\ \exp\left\{-i\frac{g_0}{\hbar}\int_\gamma A_\mu^k(y) T_k\, dy^\mu\right\}, \tag{18}$$

with P denoting path ordering, and γ is a curve in space–time that is congruent to $\hat{\gamma}$ while $\bar{\gamma}$ is the curve γ traversed in the reverse order. In (18), γ begins at x and ends at $x + \ell$, where ℓ^μ is a fixed vector (independent of x^μ), and $\bar{\gamma}$ therefore begins at $x + \ell$ and ends in x. Then

$$P\ \exp\left(-\frac{i}{\hbar}\int_{\bar{\gamma}} \hat{p}_\mu\, dy^\mu\right) = \left\{P\ \exp\left(-\frac{i}{\hbar}\int_\gamma \hat{p}_\mu\, dy^\mu\right)\right\}^\dagger. \tag{19}$$

Under a local gauge transformation, (18) transforms to

$$\begin{aligned} g'_\gamma(x) &\equiv P\ \exp\left(-\frac{i}{\hbar}\int_{\bar{\gamma}} \hat{p}'_\mu\, dy^\mu\right) P\ \exp\left\{-i\frac{g_0}{\hbar}\int_\gamma A'^k_\mu(y) T_k\, dy^\mu\right\} \\ &= P\ \exp\left(-\frac{i}{\hbar}\int_{\bar{\gamma}} \hat{p}'_\mu\, dy^\mu\right) u(x+\ell) \\ &\quad \times P\ \exp\left\{-i\frac{g_0}{\hbar}\int_\gamma A_\mu^k(y^\mu) T_k\, dy^\mu\right\} u^\dagger(x), \end{aligned} \tag{20}$$

where $\hat{p}'_0 = H'$, $\hat{p}'_i = \hat{p}_i = i\hbar\frac{\partial}{\partial x^i}$. Now write $P\ \exp(-\frac{i}{\hbar}\int_{\bar{\gamma}} \hat{p}'_\mu\, dy^\mu)$ as a product of infinitesimal exponentials of the form $(1 - \frac{i}{\hbar} p_j\, dy^j)(1 - \frac{i}{\hbar} p'_0\, dy^0)$, and use (10) and (16) in their infinitesimal forms. Then (20) implies

$$\begin{aligned} g'_\gamma(x) &= u(x) P\ \exp\left(-\frac{i}{\hbar}\int_{\bar{\gamma}} \hat{p}_\mu\, dy^\mu\right) P\ \exp\left\{-i\frac{g_0}{\hbar}\int_\gamma A_\mu^k(y) T_k\, dy^\mu\right\} u^\dagger(x) \\ &= u(x) g_\gamma(x) u^\dagger(x). \end{aligned} \tag{21}$$

Hence, the operator g_γ acts gauge covariantly on the Hilbert space. It follows from this proof that g_γ would be gauge-covariant also when $\bar{\gamma}$ in (18) is replaced by any piecewise-differentiable curve that would connect $x + \ell$ to x.

Also, on using (19),

$$\langle\psi|g_\gamma|\psi\rangle = \left\langle P\ \exp\left(-\frac{i}{\hbar}\int_\gamma \hat{p}_\mu\, dy^\mu\right)\psi \,\middle|\, P\ \exp\left(-ig_0\int_\gamma A^k_\mu T_k\, dy^\mu\right)|\psi\right\rangle$$
$$= \int d^3x\ \psi^\dagger_\gamma(x+\ell)P\ \exp\left(-ig_0\int_\gamma A^k_\mu T_k\, dy^\mu\right)\psi(x), \tag{22}$$

where $\psi_\gamma(x+\ell) \equiv P\ \exp(-\frac{i}{\hbar}\int_\gamma \hat{p}_\mu\, dy^\mu)\psi(x)$. The expectation value (22) is gauge-invariant because the integrand is gauge-invariant. It may be observable, in principle, by the Josephson effect for a non-Abelian gauge theory proposed in (Anandan, 1986). More experimental consequences of g_γ will be discussed in Section 5.

The energy–momentum operator $\hat{p}_\mu$ may be generalized to relativistic quantum field theory by defining its components as the conserved quantities obtained via Noether's theorem from the invariance of the Lagrangian under space–time translations:

$$\hat{p}_\mu = \int d^3x\ \hat{T}_{\mu 0}, \tag{23}$$

where $\hat{T}_{\mu\nu}$ is the conserved, normal-ordered energy–momentum tensor. Then the canonical commutation relations imply that $\hat{p}_\mu$ generates space–time translations. It therefore follows from arguments analogous to the above that g_γ, given by (18) with the quantum field theoretic $\hat{p}_\mu$, is gauge-covariant. Since $\hat{p}_\mu$ transforms as a covariant vector under Lorentz transformations, it also follows that g_γ is Lorentz-invariant. Moreover, g_γ may be generalized to curves γ in $\mathcal{R}^n$ (n is any positive integer), which would correspond to an n-dimensional space–time.

It may be reasonable to take g_γ as a quantum distance that replaces the classical space–time distance along the curve γ. But to do so it would be necessary to obtain the classical distance in an appropriate limit from g_γ, which will be studied in the next section.

4. CLASSICAL LIMIT

To take the classical limit of this geometry, note that classical space–time is constructed with measuring instruments consisting of particles that have approximate position and momentum. It is therefore reasonable to represent them by Gaussian wave packets, which have minimum uncertainty. For a particle with mean position at the origin and mean momentum zero, the normalized wave function of such a state up to an arbitrary phase factor is

$$\psi_{\mathbf{0}}(\mathbf{x}) = (2\pi\,\Delta x^2)^{-1/4}\ \exp\left(-\frac{\mathbf{x}^2}{4\Delta x^2}\right), \tag{24}$$

where Δx is the uncertainty in position. This may be a state of a molecule in the cup, mentioned in Section 3, in a harmonic oscillator potential in which case it would not spread. As the cup is displaced, the above wave function becomes

$$\psi_\ell(\mathbf{x}) \equiv \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\boldsymbol{\ell}\right)\psi_{\mathbf{0}}(\mathbf{x}) = (2\pi\,\Delta x^2)^{-1/4}\,\exp\left(-\frac{(\mathbf{x}-\boldsymbol{\ell})^2}{4\Delta x^2}\right) \tag{25}$$

up to a phase factor.

In (8) the second factor may be made the identity by choosing a gauge in which $\mathbf{A}\cdot\boldsymbol{\ell} = 0$. Therefore, $\exp(-i\hat{\mathbf{p}}\cdot\ell)$ may be regarded as a special case of (8) or (18), and hence as defining a quantum distance between the states $\psi_{\mathbf{0}}$ and ψ_ℓ. We may therefore expect a metric to be defined on the translation group to which this operator belongs and use that to define a metric in space. But this group, being Abelian, has no natural metric on it. There are two ways, however, that a metric may be defined on it. One is to use the Casimir operator of the Poincare group, $\eta_{ab}P^aP^b$, to define a metric on it, which locally may be associated with the space–time metric (Anandan, 1980). The other method, which will be used here, is to utilize the overlap of the two wave functions to obtain a measure of the displacement between them, which would then give an equivalent metric in the translation group. This is possible if the space of wavefunctions on which this group acts has an inner product, which would give a measure of the overlap and therefore how far a state has been translated.

Such a measure is given by the Fubini–Study metric in the quantum state space, or the set of rays, of every Hilbert space. This is the unique metric, up to multiplication by an overall constant, that is invariant under unitary (and anti-unitary) transformations. This may therefore be written in the form (Anandan, 1990, 1991; Anandan and Aharonov, 1990; Provost and Vallee, 1980)

$$dS^2 = 4(1 - |\langle\psi \mid \psi'\rangle|^2), \tag{26}$$

where dS is the infinitesimal distance between two neighboring states (rays) represented by normalized state vectors ψ and ψ'. Clearly, dS is zero when the states are the same, and it increases when the overlap between the states decreases. It is also invariant under unitary transformations, and must therefore be the Fubini–Study metric. The factor 4 in (26) is just a convention which ensures that this metric in the state space of the Hilbert subspace spanned by ψ and ψ' is the metric on a sphere of unit radius.

Now substitute $\psi_\ell(\mathbf{x})$ and $\psi_{\ell+d\ell}(\mathbf{x})$ as ψ and ψ' in (26). Then,

$$dS^2 = \frac{d\boldsymbol{\ell}^2}{\Delta x^2}, \tag{27}$$

neglecting higher order terms in $d\boldsymbol{\ell}$ because it is infinitesimal. Hence $d\boldsymbol{\ell}^2$, which is the same for all Hilbert spaces, may be used as a metric on the three-dimensional

translational group parametrized by the components of the vector ℓ. Locally, this metric may be regarded as a metric in the physical space of classical physics.

This result may be generalized to an arbitrary state ψ as follows. Require that $\langle\psi \mid \psi'\rangle$ be close to the identity, where

$$\psi'(\mathbf{x}) \equiv \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot d\boldsymbol{\ell}\right)\psi(\mathbf{x}). \tag{28}$$

Therefore expand

$$\begin{aligned}\langle\psi \mid \psi'\rangle &\simeq \left\langle\psi\left|\left(1-\frac{i}{\hbar}\hat{p}\,d\ell-\frac{1}{2\hbar^2}\hat{p}^2\,d\ell^2\right)\psi\right.\right\rangle \\ &= 1-\frac{i}{\hbar}\langle\psi|p|\psi\rangle d\ell-\frac{1}{2\hbar^2}\langle\psi|p^2|\psi\rangle d\ell^2,\end{aligned} \tag{29}$$

where $\hat{p}$ is the momentum component in the direction of $d\ell$. Substituting this in (26),

$$dS^2=\frac{4\Delta p^2}{\hbar^2}\,d\ell^2, \tag{30}$$

where Δp is the uncertainty in p: $\Delta p^2=\langle\psi|p^2|\psi\rangle-\langle\psi|p|\psi\rangle^2$. For the Gaussian wave packet, $\Delta p\Delta x=\hbar/2$, and therefore (30) gives (27) in this case.

Time is measured by a clock. Since the clock must have moving parts, the uncertainty ΔE of its Hamiltonian $\hat{H}$ must be nonzero. Neglecting any external interaction of the clock, $\hat{H}$ is a constant. The infinitesimal time evolution of the the clock is given by

$$|\psi(t+dt)\rangle=\exp\left(-\frac{i}{\hbar}\hat{H}\,dt\right)|\psi(t)\rangle. \tag{31}$$

The Fubini–Study distance dS along the evolution curve in the quantum state space corresponding to $|\psi(t)\rangle$ is obtained analogous to the derivation of (30) to be (Anandan, 1990, 1991; Anandan and Aharonov, 1990)

$$dS^2=\frac{4\Delta E^2}{\hbar^2}\,dt^2. \tag{32}$$

A quantum clock directly measures the Fubini–Study ditance S and the time t is then inferred from S using (32). The appearance of the same t in (32) in all Hilbert spaces is due to the universality of the action of the time translation $\exp(\frac{i}{\hbar}\hat{H}\,dt)$ in every Hilert space. This is analogous to the universality of the spatial displacements, which was used earlier to obtain the spatial metric.

The spatial translation (28) and the active time translation corresponding to (31) may be written covariantly as

$$\psi'=\exp\left(\frac{i}{\hbar}\hat{p}_\mu\,d\ell^\mu\right)\psi_0, \tag{33}$$

where $d\ell^\mu = (cdt, d\boldsymbol{\ell})$. This transformation is the infinitesimal version of (18) in a special gauge that makes the second factor in (18) the identity.[3] In an arbitrary gauge, the same space and time metrics are obtained by replacing the operator in (33) with the infinitesimal version of (18) in the above treatment. In relativistic quantum theory, owing to the transformation property of $\hat{p}_\mu$ mentioned at the end of Section 3, these metrics give a space–time metric that is invariant under Lorentz transformations. In the presence of gravity, this Lorentzian metric may be obtained locally by doing the above space–time translations in a freely falling Einstein elevator, which then globally gives a curved pseudo-Riemannian metric.

5. INTERACTIONS AND THE NONLOCALITY OF QUANTUM THEORY

If the quantum geometry is determined by relations between states that are group elements, and if these group elements, which are our observables, are made dynamical the way Einstein made space–time distances dynamical in order to obtain gravity, then this would give both gravity and gauge fields (Anandan, 1999). Also, quantum mechanics has an inherent nonlocality, which may also be understood as due to these group elements being the basic observables. The combination of these two statements imply that gauge fields and gravity should affect quantum states in a nonlocal manner as in the Aharonov–Bohm effect (Aharonov and Bohm, 1959), as will be discussed later.

First consider the nonlocality of quantum theory, which may be illustrated by the following example: Electrons with initial momentum in the x direction go through an infinite diffraction grating in the yz plane of a Cartesian coordinate system, with the length of the slits along the z direction. Then the grating destroys continuous translational symmetry for the electrons in the y direction. However, if the distance between successive slits in the y direction is ℓ then $s = \exp(-i\frac{p\ell}{\hbar})$ satisfies (1), where p is the momentum operator for electrons in the y direction, which generates translations in the y direction. (For simplicity, here and henceforth, the circumflex ($\hat{}$) over operators is omitted). Hence, it follows from the generalized Noether's theorem in Section 2 that $\exp(-i\frac{p\ell}{\hbar})$ is conserved although p is *not* conserved. Indeed, it is well known that the interference fringes on a screen that is parallel to and far away from the yz plane is given by $\ell \sin\theta_n = n\lambda$, where λ is the wave length and n is an integer. Therefore, the possible values of the momentum for an electron in the y direction after the interaction are $p_n = \frac{h}{\lambda}\sin\theta_n = n\frac{h}{\ell}$, i.e. $\exp(-i\frac{p_n\ell}{\hbar}) = 1$. Hence, $\exp(-i\frac{p\ell}{\hbar})$ is conserved during the passage of electrons through the grating.

The above operator s is equivalent to the modular momentum $p(\mathrm{mod}\frac{h}{\ell})$ introduced by Aharonov *et al.* (1969) on the basis of the above example. But

[3] If $d\ell^\mu$ is along γ then $dy^\mu = -d\ell^\mu$ because $\bar{\gamma}$ is the reversal of γ.

here I shall treat s as an element of a universal group that is used to define a quantum geometry as in Section 3. s may be obtained from experiments by measuring the Hermitian observables

$$s_R \equiv \frac{1}{2}\left[\exp\left(-i\frac{p\ell}{\hbar}\right) + \exp\left(i\frac{p\ell}{\hbar}\right)\right],$$
$$s_I \equiv \frac{1}{2i}\left[\exp\left(-i\frac{p\ell}{\hbar}\right) - \exp\left(i\frac{p\ell}{\hbar}\right)\right]. \tag{34}$$

Therefore, the unitary operator s may also be regarded as an observable. It is important to note that this is a *nonlocal* observable, unlike p.

This nonlocality in quantum mechanics may also be illustrated in the simple interference experiment of two coherent wave packets. Suppose that the two wave packets are moving in the x direction and have no overlap at time t. For simplicity, assume that they are the same except that their centers are separated by a displacement ℓ in the y direction. Let α be the phase difference between the wave packets. The wave is then a superposition of these two wave packets:

$$\psi(x, y, z, t) = \frac{1}{\sqrt{2}}\{\phi(x, y-\ell, z, t) + e^{i\alpha}\phi(x, y, z, t)\}. \tag{35}$$

Now no local experiments performed on the two wave packets at the two slits could determine the phase factor $e^{i\alpha}$. For example, the expectation values of the local variables p^n, where n is any positive integer, give no information about $e^{i\alpha}$ (Aharonov *et al.*, 1969). This is easily verified by writing $p^n = (-i\hbar\frac{\partial}{\partial y})^n$ in the coordinate representation. But

$$\langle\psi|\exp\left(-i\frac{p\ell}{\hbar}\right)|\psi\rangle = \frac{e^{i\alpha}}{2}. \tag{36}$$

This means that the momentum distribution at time t does depend on the phase factor $e^{i\alpha}$, i.e. if p is measured then the probability distribution for obtaining the individual eigenvalues of p is changed by this phase factor (but the average $\langle p\rangle$ is unchanged). And this may be experimentally verified by letting the wave packets interfere and observing the shift in interference fringes. Hence, $\langle\psi|\exp(-i\frac{p\ell}{\hbar})|\psi\rangle$ contains more information than the expectation values $\langle\psi|p^n|\psi\rangle$ of any of the moments of momentum p^n. This is basically due to the linear structure of the Hilbert space, which physically corresponds to the principle of superposition, and the fact that ψ is not an analytic function.

This fundamental nonlocality of quantum mechanics translates into a nonlocality of the effect of all the fundamental interactions on the wave function. This has been shown for the Aharonov–Bohm effect due to a magnetic field by Aharonov *et al.* (1969). It may be illustrated in the above described interference of two wave packets as follows. Suppose the above two wave packets A and B

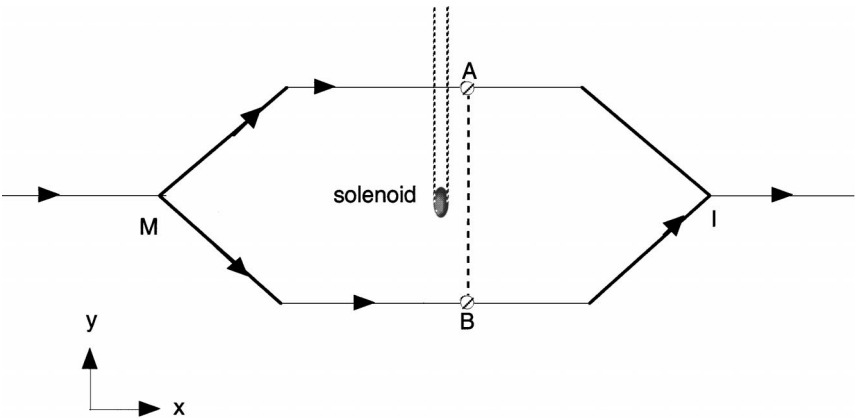

Fig. 1. Vector Aharonov–Bohm effect in which a wave packet of an electron is split coherently into two wave packets at the beam splitter M and then made to interfere at I. When the imaginary line joining the centers of the wave packets A and B sweeps across the magnetic flux in the solenoid (shaded region) the modular momentum or the modular kinetic momentum associated with this line changes, as pointed out by Aharonov (private communication).

are those of an electron and they pass on the two sides of a solenoid containing a magnetic flux Φ. The gauge may be chosen so that the vector potential is nonzero only along a thin strip bounded by two planes indicated by the dotted lines in Fig. 1. Then when the line AB passes the solenoid, the wave packet A acquires a phase difference $\alpha = \frac{e}{\hbar c}\Phi$ with respect to the wave packet B. Therefore, the expectation value of the modular momentum s, given by (36), which was $\frac{1}{2}$ before the line AB passed the solenoid is now $\exp(i\frac{e}{\hbar c}\Phi)/2$. It was pointed out by Aharonov (private communication) that in the last statement s may be replaced by the gauge-invariant modular kinetic momentum f_ℓ given by (8). This is because before and after the wave packets pass the solenoid the vector potential $\mathbf{A}$ is zero along the line AB, and therefore f_ℓ is the same as s. Since f_ℓ is gauge-covariant, $\langle\psi|\, f_\ell|\psi\rangle$ is gauge-invariant. Hence, the above-mentioned change in $\langle\psi|\, f_\ell|\psi\rangle$ by the factor $\exp(i\frac{e}{\hbar c}\Phi)$, as the line AB crosses the solenoid, is the same in every gauge.

However, we may replace f_ℓ in the above arguments by the more general gauge-covariant operator g_γ given by (18) with γ here being an arbitrary (piecewise-differentiable) spacelike curve that joins A and B. Then

$$g_\gamma(x) = \exp\left(-\frac{i}{\hbar}\hat{\mathbf{p}}\cdot\ell\right)\exp\left\{i\frac{e}{\hbar}\int_{\gamma\mathbf{x}}^{\mathbf{x}+\ell}\mathbf{A}(\mathbf{y},t)\cdot d\mathbf{y}\right\}, \tag{37}$$

where the integral is from $\mathbf{x}$ to $\mathbf{x}+\ell$ along γ. Hence,

$$\langle\psi|g_\gamma|\psi\rangle = \int d^3x\ \psi^\dagger(\mathbf{x}+\ell,t)\exp\left\{i\frac{e}{\hbar}\int_{\gamma\mathbf{x}}^{\mathbf{x}+\ell}\mathbf{A}(\mathbf{y},t)\cdot d\mathbf{y}\right\}\psi(\mathbf{x},t), \tag{38}$$

which is gauge-invariant. Then $\langle\psi|g_\gamma|\psi\rangle$ changes by the factor $e^{\pm i\frac{e}{\hbar c}\Phi}$ as a portion of γ crosses the solenoid. It follows that there is nothing special about the straight line AB crossing the solenoid in the experiment described in Fig. 1. The arbitrariness of the choice of γ joining A and B in the above argument reflects the topological nature of the Aharonov–Bohm effect. What is ultimately observed in this effect is the phase factor $\exp(-i\frac{e}{\hbar}\oint A_\mu\, dx^\mu)$ where the integral is around the solenoid or more generally around the region in which the field strength is nonvanishing. Since this is an integral of the 1-form A_μ along a curve, it contains no information about the metric of space–time.[4] To determine the straight line, or a geodesic in general, the metric is needed, and the above phase factor therefore cannot show any preference to a geodesic such as the line AB. The integrand of (38) (with e replaced by $2e$) was previously used in Ref. (Anandan, 1986) to study the Josephson effect due to the enclosed magnetic flux in a superconducting ring that has a Josephson junction, which is also an Aharonov–Bohm effect.

Consider now the scalar Aharonov–Bohm effect. A wave packet traveling in the x direction is partially transmitted and reflected by a beam splitter M (Fig. 2(a)). The two resulting wave packets, which travel in opposite directions, are reflected by two mirrors M_1 and M_2 situated along the x axis and they interfere subsequently. Meanwhile a pair of oppositely charged capacitor plates C is separated and closed so that there is a nonzero electric field in the region enclosed by the worldlines of the centers of the wave packets in the xt plane. The same experiment is viewed in the rest frame of the reflected wave packet in Fig. 2(b).

We may choose a gauge in which the vector potential is nonzero only along a strip between the dotted lines parallel to the time axis in Fig. 2(a) or 2(b). Then the wave packet A at time $t+\tau$ develops a phase shift β with respect to B at time t as the imaginary line AB crosses the space–time region containing the electric field, where

$$\beta = \frac{e}{\hbar}\int_C F_{0x}\, dx\, dt \tag{39}$$

and C is the region in which the electric field $E = F_{0x}$ is nonzero. This is a nonlocal effect, which may be understood using $U = \{T\ \exp(-\frac{i}{\hbar}\int_0^\tau H\, dt)\}^\dagger$, where T denotes time ordering and τ is the time interval between the events A_1 and B_1. We shall assume that the time dependence of H comes only from the vector potential contained in H. Suppose the wave function of the electron is

$$\psi = \frac{1}{\sqrt{2}}(\psi_A + \psi_B),$$

[4] This explains why in the nonrelativistic limit for the charged particle, the Ahaonov–Bohm phase shift that it experiences due to the phase factor $\exp(-i\frac{e}{\hbar}\oint A_\mu dx^\mu)$ remains the same. This is because the latter phase factor is unaffected by the change in the space–time metric that takes place in the nonrelativistic limit.

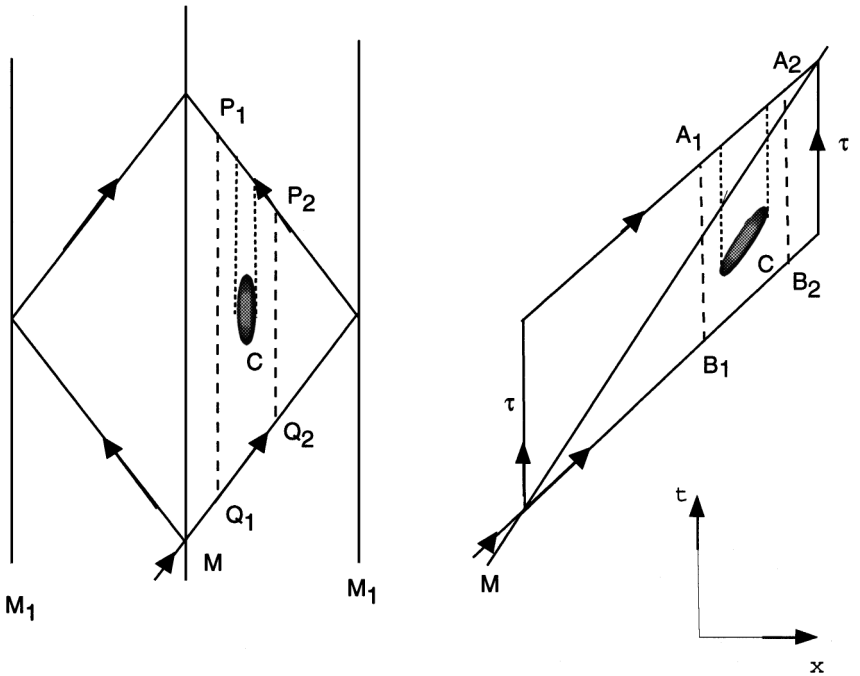

Fig. 2. A scalar Aharonov–Bohm effect shown schematically in the *tx* plane. (a) The modular energy and modular kinetic energy associated with the imaginary lines P_1Q_1 and P_2Q_2 are different, partly because of the scalar AB phase shift due to the region of nonzero electric field in the capacitor that is open and shut (shaded region). (b) The same experiment viewed in the rest frame of the wave packet that is reflected at the beam splitter M. When the imaginary line AB joining the wave packets sweeps across the region of nonzero electric field in the capacitor the modular energy and modular kinetic energy associated with this line changes because of the AB phase shift.

where ψ_A and ψ_B are the wave functions of the above two localized wave packets. Then

$$\langle\psi(t)|U|\psi(t)\rangle = \frac{1}{2}\langle\psi_A(t)|U|\psi_B(t)\rangle = \frac{1}{2}\langle\psi_A(t+\tau)\mid\psi_B(t)\rangle \qquad (40)$$

changes by the factor $e^{i\beta}$ due to the electric field E in the region C as the imaginary line AB sweeps across the small region C where E is nonzero (Fig. 2(b)).

In the last statement and in (40) we may replace U by the gauge-covariant modular kinetic energy operator f_τ given by (14), with e replacing q. This is because A_0 is zero along A_1B_1 and A_2B_2, which are outside the strip that contains the nonzero A_μ. However, $\langle\psi(t)|\,f_\tau\,|\psi(t)\rangle$ is gauge-invariant and therefore its change

mentioned above may be obtained in any gauge. But since there are no forces acting on the electron, there is no change of its kinetic energy $H - eA_0$ or any of its moments. Hence, $\langle\psi(t)|\, f_\tau|\psi(t)\rangle$ contains nonlocal aspects that cannot be obtained by measurements of the kinetic energy operator or any of its moments. Thus the scalar Aharonov–Bohm effect may be viewed as a quantum effect that is *nonlocal in time.*

The above results are easily generalized to the Aharonov–Bohm effects due to non-Abelian gauge fields (Anandan, 1979) by replacing the electromagnetic fluxes by fluxes of Yang–Mills field strength, and using the expectation values (22) of the gauge-covariant operator g_γ given by (18). In all cases, the gauge-invariant $\langle\psi|g_\gamma|\psi\rangle$ changes as γ passes a cross-section of the gauge field flux. If the gauge field flux has a singular cross-section, then γ passes the flux at an event, which of course is independent of the inertial frame in which the effect is being described, but depends on the choice of γ. When γ passes this event, in this idealized case, $\langle\psi|g_\gamma|\psi\rangle$ changes to $\langle\psi|g_\gamma P\ \exp\{-i\frac{g_0}{\hbar}\oint_C A^k_\mu(y)T_k\, dy^\mu\}|\psi\rangle$, where C is a closed curve going around the Yang–Mills flux.

The group element (18) belongs to the group $T_4 \times G$, where for a closed system T_4 is the translation group and G is the gauge group. But it has an asymmetry in that the part of (18) that belongs to G is dynamical, whereas the part that belongs to T_4 is fixed. Since I proposed that the fundamental interactions should correspond to the universal group element (18) being dynamical, consistency requires that the part of (18) that belongs to T_4 should be dynamical as well, i.e. ℓ^μ should be made dynamical. But the classical spacetime geometry was constructed in Section 4 using the latter group elements. It follows therefore that making ℓ^μ dynamical would make the space–time metric dynamical and not fixed as it is in Minkowski space–time. Therefore, the interaction that corresponds to making the T_4 group elements dynamical gives, in the classical limit, the well-known geometrical description of gravity in classical general relativity. A generalization of it is obtained by replacing T_4 with T_n, where n is any positive integer. Thus the present approach requires the existence of gravity.

I now give a simple illustration of the above unified way of treating gravity and gauge fields by considering the gravitational analog of the above vector Aharonov–Bohm effect. The geometry surrounding a cosmic string in the two dimensional section normal to the axis of the string at a given time is that of a cone whose center is at the axis, which is seen by solving the classical gravitational field equations (Anandan, 1996). The space–time geometry of a nonrotating cosmic string is obtained by simply adding to this plane the extra dimension in the direction of the axis and the time dimension; then the curvature outside the string is zero everywhere (Fig. 3). It is known that this geometry is similar to the electromagnetic field around a solenoid because the curvature is zero outside the string and yet there is a nontrivial holonomy around it.

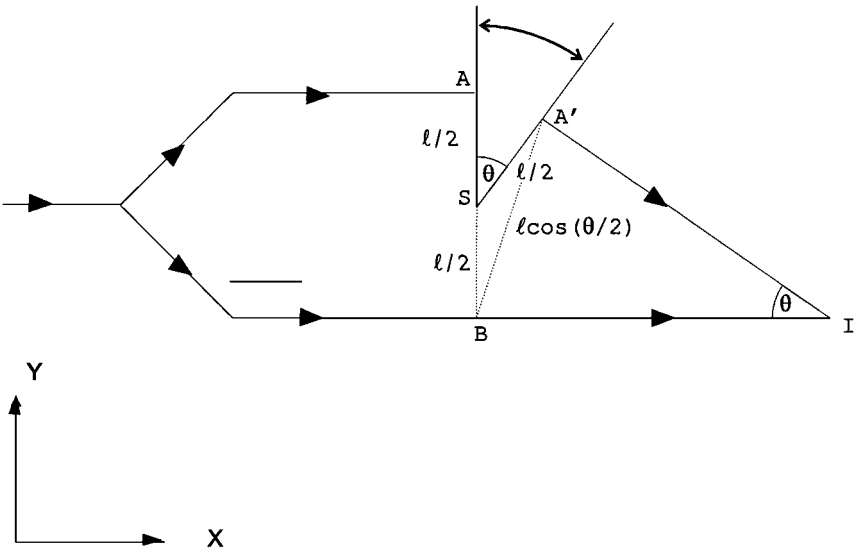

Fig. 3. A gravitational analog of the experiment in Fig. 1. The conical geometry around a cosmic string that is normal to the plane through S is represented by cutting off the wedge ASA' from flat space and identifying the planes along which it is cut. The wave packets moving at A (same as A') and B are focused by this geometry to interfere at I. Just before AB crosses S, there are two geodesics connecting A and B of lengths ℓ and $\ell\cos(\theta/2)$ in the curvature-free region. But just after the crossing there is a unique geodesic $A'B$ of length $\ell\cos(\theta/2)$ joining A and B.

Consider now two wave packets separated by a distance ℓ and whose centers move along initially parallel lines such that the geodesic line AB joining the centers meets the conical singularity S at its midpoint. But there is another geodesic that connects the same pair of points A and B of length $\ell\ \cos(\theta/2)$, shown by the line $A'B$ in Fig. 3, where A' is identified with A. (Actually, there are two geodesics connecting A and B when the angle $A\hat{S}B$ on the left side exceeds $\pi - \theta$ but is less than π. The corresponding angles $A\hat{S}B$ on the right side are π and $\pi - \theta$, respectively. The singularity may be replaced by a small smooth cap that merges with the rest of the cone smoothly. Then there would be a third geodesic joining A and B through this cap.)

Just after AB crosses the conical singularity, there is only one geodesic joining A and B, whose length is $\ell\ \cos(\theta/2)$. This is somewhat analogous to the change in modular kinetic momentum when AB crosses the solenoid in Fig. 1, because $\exp(-i\mathbf{p}\cdot\boldsymbol{\ell})$ now translates locally along parallel geodesic line segments of length ℓ, and this length has gotten shorter after the crossing.

In Fig. 4, this result is generalized to the case of S not being the midpoint of the geodesic ASB, and it is also seen to be "gauge" independent in the sense of

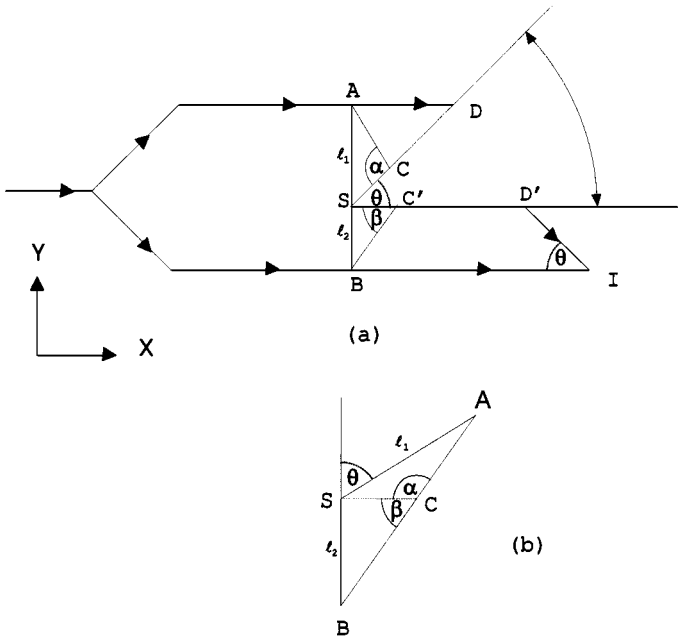

Fig. 4. A generalization of the gedanken experiment shown schematically in Fig. 3 to unequal lengths $AS = \ell_1$ and $BS = \ell_2$. The wedge *DSD′* has an arbitrary orientation. But since *SC* and *SC′* are identified, and $\alpha + \beta = \pi$ in order for *ACB* to be a geodesic, the distances along the geodesics *ASB* and *ACB* are respectively $\ell_1 + \ell_2$ and $(\ell_1^2 + \ell_2^2 + 2\ell_1\ell_2 \cos\theta)^{1/2}$, as shown in (b), independently of the orientation of the wedge.

being invariant under the rotation of the "wedge" mentioned above. There is then a phase shift $\Delta\phi$ due to the difference in path lengths traveled by the wave packets given by

$$\Delta\phi = \frac{p_0}{\hbar}(d_1 - d_2) \simeq \frac{p_0}{2\hbar}(\ell_2 - \ell_1)\theta = 4\pi G\mu\frac{p_0}{\hbar}(\ell_2 - \ell_1) \tag{41}$$

for small θ, where d_1 and d_2 are the path lengths AI and BI, μ is the mass per unit length of the cosmic string, and p_0 is the initial momentum of the particle. If the particle carries spin, then there is also a phase shift because of the coupling of spin to the curvature.

This and other phase shifts for interference of two wave packets around a cosmic string are studied elsewhere and are understood as being due to the Poincare holonomy around the string (Anandan, 1994, 1996). In the present approach, these phase shifts may be understood by associating with each curve an element of the

Poincare group element (Anandan, 1999) whose expectation value changes as the curve crosses the cosmic string.

6. DISCUSSION AND CONCLUSION

The change in geodesic distances between the wave packets due to the cosmic string, in Section 5, is not surprising because gravity changes distances, according to general relativity, and the cosmic string is a purely general relativistic object without a Newtonian analog. What may be more interesting is the similar change of the "quantum distances" due to the electromagnetic field in the usual Aharonov–Bohm effect and its generalization to non-Abelian gauge fields. Both these effects may be treated in a somewhat analogous manner if the modular kinetic energy–momentum (18), regarded here as universal group elements, may be interpreted as "distances" in a quantum geometry as proposed earlier. The fact that space–time distances may be obtained from them approximately, as shown in Section 4, reinforces this view.

The treatment of (18) as an observable, which implies the above-mentioned nonlocal effects in quantum theory, perhaps removes the mystery of why although the interactions are local as they occur in the Hamiltonian or Lagrangian, there are nevertheless nonlocal effects such as the Aharonov–Bohm and its generalizations to non-Abelian gauge fields and gravitation.[5]

A criticism that may be made against regarding the universal group elements (18) as quantum distances is that their action on wave functions seems to require that $\hat{\gamma}$ be interpreted as a set of space–time curves, whereas it was argued in Section 3 that space–time geometry is not appropriate for quantum theory. However, as discussed in that section, this limitation becomes critical only at the Planck scales. But at the Planck scale these group elements need not be associated with curves in space–time. They may be defined simply as operators acting on quantum states defining the quantum geometry and representing the fundamental interactions. Thus it is possible to have a quantum geometry even at scales in which the space–time geometry breaks down and this would also give a quantum description of all the interactions.

ACKNOWLEDGMENTS

I thank Yakir Aharonov, Alonso Botero, and Ralph Howard for useful discussions. This work was partially supported by NSF Grant 9971005 and ONR Grant N00014-00-1-0383.

[5] This reconciles the views of Y. Aharonov and C. N. Yang, who regarded the Aharonov–Bohm effect as being nonlocal and local, respectively, in *Proceedings of the International Symposium on the Foundations of Quantum Mechanics*, Tokyo, August 1983, edited by S. Kamefuchi *et al.* (Physical Society of Japan, Tokyo, 1984, pp. 65–73.

REFERENCES

Aharonov, Y. and Bohm, D. (1959). *Physical Review* **115**, 485.
Aharonov, Y., Pendleton, H., and Peterson, A. (1969). *International Journal of Theoretical Physics* **2**, 213.
Anandan, J. (1979). *Il Nuovo Cimento* **53A**, 221.
Anandan, J. (1980). *Foundations of Physics* **10**, 601–629.
Anandan, J. (1986). *Physical Review D* **33**, 2280–2287.
Anandan, J. (1990). *Physics Letters A* **147**, 3.
Anandan, J. (1991). *Foundations of Physics* **21**, 1265.
Anandan, J. (1994). *Physics Letters A* **195**, 284.
Anandan, J. (1996). *Physical Review D* **53**, 779. gr-qc/9507049, and references therein.
Anandan, J. (1999). *Foundations of Physics* **29** (11), 1647–1672. quant-ph/9808045.
Anandan, J. and Aharonov, Y. (1990). *Physical Review Letters* **65**, 1697–1700.
Kibble, T. W. B. (1979). *Communications in Mathematical Physics* **65**, 189.
Provost, J. P. and Vallee, G. (1980). *Communications in Mathematical Physics* **76**, 289.
Synge, J. L. (1960). *Relativity: The General Theory*, North Holland, Amsterdam, Chap. III.
Wigner, E. (1967). *Symmetries and Reflections*, Indiana U. P., Bloomington.
Wisnievesky, D. and Aharonov, Y. (1967). *Annals of Physics* **45**, 479.
Yang, C. N. (1970). *Physical Review D* **1**, 2360.